

UNIVERSIDADE FEDERAL DO PARANÁ

REGIANE ABJAUD ESTOPA

FENOTIPAGEM VIA NIR E PREDIÇÃO GENÔMICA EM *Eucalyptus benthamii*

Curitiba – PR

2017

REGIANE ABJAUD ESTOPA

FENOTIPAGEM VIA NIR E PREDIÇÃO GENÔMICA EM *Eucalyptus benthamii*

Tese apresentada ao Curso de Pós-Graduação em Agronomia, Área de Concentração em Produção Vegetal, Departamento de Fitotecnia e Fitossanitarismo, Setor de Ciências Agrárias, Universidade Federal do Paraná, como parte das exigências para obtenção do título de Doutora em Ciências.

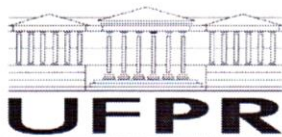
Orientador: Prof. Dr. Ricardo Augusto de Oliveira

Coorientador: Prof. Dr. Marcos Deon Vilela de Resende

Coorientadora: Prof. Dra. Camila Ferreira Azevedo

Curitiba – PR

2017

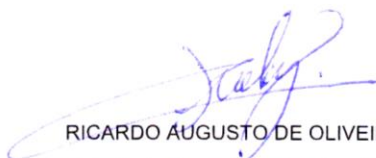


MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
Setor CIÊNCIAS AGRÁRIAS
Programa de Pós-Graduação AGRONOMIA (PRODUÇÃO VEGETAL)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em AGRONOMIA (PRODUÇÃO VEGETAL) da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **REGIANE ABJAUD ESTOPA** intitulada: **FENOTIPAGEM VIA NIR E PREDIÇÃO GENÔMICA EM *Eucalyptus benthamii***, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO.

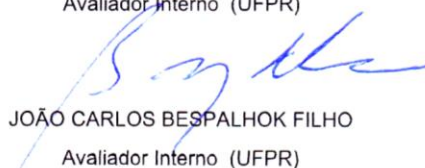
Curitiba, 16 de Março de 2017.



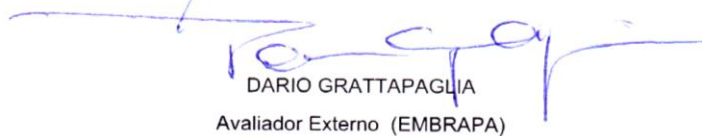
RICARDO AUGUSTO DE OLIVEIRA
Presidente da Banca Examinadora (UFPR)



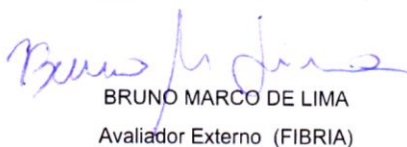
BRUNO PORTELA BRASILEIRO
Avaliador Interno (UFPR)



JOÃO CARLOS BESPALHOK FILHO
Avaliador Interno (UFPR)



DARIO GRATTAPAGLIA
Avaliador Externo (EMBRAPA)



BRUNO MARCO DE LIMA
Avaliador Externo (FIBRIA)

OFEREÇO

A Deus que sempre me guia para o caminho do bem e do sucesso.
Aos meus pais que sempre me apoiam e ajudam na caminhada da vida.

DEDICO

Ao meu esposo pelo companheirismo, amizade e amor.
A minha família pelo apoio e amor.

AGRADECIMENTOS

Primeiramente a Deus por ter me dado força e motivação para realizar o doutorado trabalhando.

A empresa Klabin que me possibilitou essa capacitação por meio do doutorado e desenvolvimento desse projeto. Ao Mario G. Ladeira, Carlos A. S. Amaral Santos, José Artemio Totti e Francisco C. Razzolini pela confiança e apoio neste projeto.

Aos meus pais, Joana Darc e Donizetti, agradeço pelo amor e carinho. Por sempre confiarem no meu potencial e me apoiarem nas minhas decisões.

As minhas irmãs Apoliana e Aline, ao meu sobrinho Aleandro, incentivo, carinho e confiança.

Ao meu esposo Celso, pois estive ao meu lado em todos os momentos difíceis e de sucesso desse doutorado. Obrigada pelo seu apoio, compreensão, amor e companheirismo.

Ao meu orientador Dr. Prof. Ricardo Augusto de Oliveira por dar significado à palavra orientador, “o que orienta, direciona; condutor, guia” sendo apoiador, incentivador e motivador desse projeto. Muito obrigada pela confiança e valorização que me acolheu. Aos professores da UFPR pelos ensinamentos.

Ao grande Deon (Marcos Deon Vilela de Resende), Dr. Professor e Pesquisador - (EMBRAPA/UFV) pelo grande apoio para que esse trabalho se concretizasse, pela coorientação, confiança e apoio de sempre.

Em especial a Professora Dra. Camila Ferreira Azevedo – UFV, pela coorientação pelo apoio e execução das análises.

Ao Dr. Dario Grattapaglia pelo apoio, incentivo e contribuições desde a concepção desse trabalho até a participação na banca de defesa. E qualidade final dos artigos produzidos.

Ao Dr. Bruno Lima por toda contribuição e participações na qualificação e defesa.

Aos professores Dr. João Carlos Bessalho Filho e ao Dr. Bruno Portela Brasileiro pelas contribuições durante a defesa.

A Dra. Flaviana Milagres pela orientação, paciência e apoio na construção dos modelos

NIR. Ao Dr. Prof. Paulo Hein pelo apoio no artigo sobre os modelos NIR.

Aos colegas e amigos da Pesquisa Florestal que sempre me incentivaram e apoiaram Fabrício, James, Leticia, Marco Figura, Mariane e Ricardo. Em especial ao Valdir que é meu braço direito e sempre me apoiou.

Aos colegas de profissão e amigos que sempre me apoiaram Teotônio de Assis, Mario Ladeira, Carla Garcia. Ao Wagner Costa e equipe que me apoiaram durante a fase árdua de coleta de baguetas na Unidade de Otacílio Costa em SC.

A querida Ana Lúcia que me acolheu na fria Curitiba toda semana durante as disciplinas, muito obrigada pelo carinho, apoio e amizade.

A família Urbano (Edilson, Fernanda e Zoe) que me acolheram durante o período do doutorado e pela amizade de sempre.

Aos meus amigos que me apoiaram e incentivaram nessa caminhada: Mara e Léo, Cida e Osmar, Andréia e Vagner, Cristiane e James e Claudinha e Mário.

Aos meus amigos e familiares que estão distantes, mas que sempre estão torcendo por mim.

SIGLAS

BLUP	Best Linear Unbiased Predictor
CTAB	Cationic hexadecyl trimethyl ammonium bromide
DAP	Diâmetro a Altura do Peito
DBM/WBD	Densidade Básica da Madeira
ESALQ/USP	Escola Superior de Agricultura Luiz de Queiroz / Universidade de São Paulo
NIR	Espectroscopia do Infravermelho Próximo
RR-BLUP	Random Regression Best Linear Unbiased Predictor
SAM/MAS	Seleção Assistida por Marcadores/ Marker Assisted Selection
SG	Seleção Genômica
SNP	Single Nucleotide Polymorfism
SRI	Seleção Recorrente Intrapopulacional
SRR	Seleção Recorrente Recíproca
UFLA	Universidade Federal de Lavras
UFPR	Universidade Federal do Paraná
UFV	Universidade Federal de Viçosa
UNESP	Universidade Estadual de São Paulo
VGG	Valores Genéticos Genômicos
VOL3	Volume aos 3 anos de idade

RESUMO

O melhoramento tradicional contribuiu por muitos anos para o aumento expressivo de produtividade do *Eucalyptus* no Brasil, porém, para completar um ciclo de melhoramento leva-se muito tempo. A seleção genômica tem apresentado potencial para acelerar os programas, e os melhoristas estão com uma expectativa grande sobre essa técnica, visto que une a biotecnologia com a genética quantitativa para acelerar a seleção genética. O objetivo principal desse trabalho foi validar biotecnologias que possam acelerar os ganhos de seleção dos programas de melhoramento genético de *Eucalyptus benthamii*. Para isso, foi avaliada a Seleção Genômica para caracteres de crescimento e qualidade da madeira. Também, foi realizada a caracterização não destrutiva da madeira por meio da espectroscopia de infravermelho próximo (NIR) para estimativa de densidade da madeira, dos teores de carboidratos totais, lignina e extrativos. Para o desenvolvimento dos modelos de seleção genômica ampla foi realizada a fenotipagem da população de ‘treinamento’ aos três anos de idade quanto ao volume de madeira, em seguida foi realizada a genotipagem utilizando marcadores moleculares do tipo SNPs em 780 indivíduos selecionados por meio de modelos mistos. Essa genotipagem foi processada utilizando um chip de alta densidade que resultou em 43.266 marcadores SNPs, após o controle de qualidade 27.462 marcas foram utilizadas para a estimativa dos efeitos. A predição de valores genéticos genômicos foi obtida para o caráter volume de madeira aos 3 de idade, e aos 4 anos para densidade da madeira, lignina, extrativos e carboidratos, utilizando o modelo RR-BLUP (Random Regression Best Linear Unbiased Predictor), sendo que as capacidades preditivas foram estimadas por diferentes metodologias de validação. A capacidade preditiva variou de 0,24 a 0,74 para volume de madeira aos três anos de idade. Para densidade básica da madeira, teor de lignina, extrativos e carboidratos, os valores de capacidade preditiva foram 0,44, 0,39, 0,33 e 0,35 respectivamente. Conclui-se que a seleção genômica ampla é aplicável para a espécie *Eucalyptus benthamii*. Com 2 a 5 mil marcadores consegue-se acessar a variabilidade para as características de crescimento e qualidade da madeira. A tecnologia NIR também se apresenta como uma potencial ferramenta para o programa de melhoramento de *E. benthamii* tornando possível a avaliação precoce e não destrutiva de características físico-químicas importantes para o processo de produção de polpa celulósica.

Palavras-chave: *Eucalyptus benthamii*, Seleção Genômica; SNPs, espectroscopia near-infrared.

ABSTRACT

Traditional breeding has contributed for many years to increase productivity of *Eucalyptus* in Brazil. However, the time to complete a breeding cycle is long and genomic selection has potential to accelerate programs. The breeders have a great expectation about this technique, because biotechnology works with quantitative genetics to accelerate the genetic selection. The main objective of this work was to evaluate biotechnologies that accelerate the genetic improvement program of *Eucalyptus benthamii*. For this, the study of the technique of genomic selection for traits of growth and wood quality was carried out. In addition, *Eucalyptus benthamii* wood was non-destructively characterized and the performance of near infrared spectroscopy (NIR) in estimating the wood basic density, lignin, extractive contents and total carbohydrates was evaluated. To develop genomic selection models was done the phenotyping of the training population with three years old for wood volume trait; the genotyped was done using SNPs molecular markers to 780 selected individuals by traditional BLUP. The genotyping was performed using a high density chip which resulted in 43,266 SNPs markers, after quality control marks 27,462 were used to estimate the effects of markers. The prediction ability was estimated by RR-BLUP model (Random Regression Best Linear Unbiased Predictor) with different validation methods by the new estimator for wood volume. The prediction model was developed at 4 years for wood density, lignin, extractives and carbohydrates, too. The prediction ability was 0,24 to 0,74 for wood volume at three years old. For wood density, lignin content, extractives and carbohydrates, the prediction ability was 0,44, 0,39, 0,33 e 0,35, respectively. In conclusion, the genome selection applies to the species of *Eucalyptus benthamii*. With 2000 to 5000 markers can be access the genetic variability of this population with good accuracy. And, the NIR technology presents itself as a potential tool for the *E. benthamii* breeding program, enabling the early and non-destructive evaluation of important physical and chemical wood features for the pulp production process and allowing cost reduction of chemical analysis, which may also be optimized.

Keywords: *Eucalyptus benthamii*, Genomic selection, SNPs, near-infrared spectroscopy.

LISTA DE FIGURAS

Figura 1	Esquema do princípio operacional de um sistema de espectrofotometria com transformada de Fourier.	30
Figura 2	Comparação da capacidade preditiva para volume de madeira aos 3 anos utilizando três métodos de validação. Validação 1- validação de 78 ciclos usando os efeitos médios estimados dos marcadores. Validação 2 - validação de 10 ciclos usando os efeitos médios estimados dos marcadores. Validação 3 - metodologia de validação de 78 ciclos usando marcadores de maiores efeitos.	55
Figura 3	Capacidade preditiva (r) para características de densidade básica da madeira, extrativos, lignina e carboidratos aos 4 anos de idade para todos os marcadores SNP, variando de mil em mil.	57
Figura 4	Demonstração dos efeitos SNPs com três mil (A) e 27.462 marcadores (B) ao longo dos onze cromossomos do genoma de <i>E. benthamii</i> para densidade da madeira.	62
Figura 5	Demonstração dos efeitos SNPs com três mil (A) e 27.462 marcadores (B) ao longo dos onze cromossomos do genoma de <i>E. benthamii</i> para teor de lignina total da madeira.	62
Figure 6	Samples being taken from the standing tree (A). Increment core (B) and saturated samples (C) for determination of basic density.	72
Figure 7	NIR predicted versus measured values plot for wood basic density (A), lignin (B), extractive (C), glucose (D), xylan (E) and total carbohydrate (F) for independent validation test set.	78

LISTA DE TABELAS

Tabela 1	Análises químicas da madeira e normas utilizadas para determinação.	44
Tabela 2	Descrição dos ciclos e grupos para a validação cruzada via procedimento Jackknife para características genotipadas e fenotipadas.	50
Tabela 3	Parâmetros genéticos da população de <i>E. benthamii</i> para características de crescimento e qualidade da madeira.	52
Tabela 4	Acurácia de seleção genômica (rgg*) calculada na população de validação associada à herdabilidade individual para o caráter volume de 0.20, considerando herdabilidade genômica (h^2M) e correlação ou capacidade preditiva (rgf), com os marcadores de maiores efeitos ordenados de 1.000 em 1.000 de forma cumulativa.	54
Tabela 5	Ponto ótimo de capacidade preditiva dos marcadores SNP e as herdabilidades, acurácias e erros associados.	58
Tabela 6	Superioridade da seleção genômica ampla (AG - Acurácia Genômica RRBLUP) comparada com a acurácia da seleção fenotípica (AF - Acurácia Fenotípica BLUP) em função da redução de tempo do melhoramento tradicional convencional em anos (TF-Tempo Fenotípica) para tempos reduzidos (TG-Tempo Genômica) para características de qualidade da madeira.	61
Table 1	Calibration of NIR prediction models for basic density (g/cm ³), total lignin (%), extractive (%), glucose (%), xylan content (%) and total of carbohydrates (%);	79
Table 2	Genetics parameters estimated for wood traits determinate by NIR models.	84
Table 3	Wood physicochemical characterization of 4 years old <i>Eucalyptus benthamii</i> population.	84

BIOGRAFIA

Regiane Abjaud Estopa, filha de Joana Darc Abjaud Estopa e Donizetti Luiz Estopa, nasceu em 12 de junho de 1978 em Nepomuceno-MG. Em 1999 ingressou no Curso de Engenharia Florestal na Universidade Federal de Lavras (UFLA), Lavras- MG, graduando-se em agosto de 2004. Nesse mesmo ano, iniciou o Mestrado em Genética e Melhoramento de Plantas (UFLA), com defesa da dissertação em junho de 2006. Em seguida ingressou no Doutorado, mas resolveu abandonar o curso e trabalhar para adquirir experiência e posteriormente retornar ao doutorado. Então em setembro de 2006, ingressou na Jari Celulose S.A., onde atuou como Pesquisadora Pleno Melhoramento Genético, Coordenadora de Pesquisa e Desenvolvimento Florestal, permanecendo na empresa até abril de 2010. Durante sua permanência na empresa, em abril de 2007, iniciou curso de Especialização em Cultura de Tecidos Vegetais na Universidade Federal de Lavras (UFLA), concluindo em abril de 2009. Em maio de 2010, ingressa na Klabin S.A. como Especialista em Melhoramento Genético. Em março de 2013 ingressa no doutorado na Universidade Federal do Paraná (UFPR). Em 2014, inicia o Curso de Pós Graduação em Celulose e Papel da Universidade Federal de Viçosa (UFV).

Atualmente é Especialista em Melhoramento Genético de *Eucalyptus* da Klabin S.A.

SUMÁRIO

1. INTRODUÇÃO GERAL	13
2. REVISÃO DE LITERATURA.....	15
2.1 <i>Eucalyptus benthamii</i> E O MELHORAMENTO GENÉTICO	15
2.3 O PROCEDIMENTO RR-BLUP GENÔMICO.....	21
2.4 IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA	25
2.5 ESTIMADOR DA ACURÁCIA DE SELEÇÃO GENÔMICA	26
2.6 A TECNOLOGIA DE ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO (NIR)	28
3. REFERÊNCIAS	32
CAPÍTULO I - CAPACIDADE PREDITIVA DO MODELO DE SELEÇÃO GENÔMICA EM <i>Eucalyptus benthamii</i> PARA VOLUME E QUALIDADE DA MADEIRA.....	38
RESUMO	38
ABSTRACT	39
1. INTRODUÇÃO.....	40
2. MATERIAL E MÉTODOS.....	43
2.1 FENOTIPAGEM.....	43
2.1.1 Fenotipagem para volume de madeira.....	43
2.2 GENOTIPAGEM	46
2.3 ANÁLISE GENÔMICA	47
3. RESULTADO E DISCUSSÃO.....	51
3.1 CARACTERIZAÇÃO DA POPULAÇÃO EM ESTUDO	51
3.2 PREDIÇÃO GENÔMICA.....	52
3.2 EFEITOS DOS MARCADORES SNP	61
4. CONCLUSÕES	63
5. REFERÊNCIAS	64
CAPÍTULO II - NIR SPECTROSCOPIC MODELS FOR WOOD TRAITS IN BREEDING PROGRAMS OF <i>Eucalyptus Benthamii</i>	68
1. INTRODUCTION	69
2. MATERIALS AND METHODS	71
2.1 GENETIC MATERIAL	71
2.2 SAMPLE COLLECTION AND PREPARATION	71
2.3 SPECTRAL READINGS	72

2.4 CHEMICAL ANALYSIS OF WOOD	72
2.5 DETERMINATION OF WOOD BASIC DENSITY	73
2.6 MODEL CALIBRATION	73
2.7 USE OF NIR MODELS IN THE BREEDING PROGRAM	75
3. RESULTS AND DISCUSSION	76
3.1 NIR SPECTRA OF WOOD	76
3.2 NIR MODELS FOR WOOD TRAITS	76
3.2.1 Wood Basic Density	76
3.2.2 Total Lignin	80
3.2.3 Extractives	81
3.2.4 Carbohydrates	82
3.3 GENETICS PARAMETERS BY NIR MODELS	83
4. CONCLUDING REMARKS	85
5. REFERENCES	86
6. CONSIDERAÇÕES FINAIS	91
7. REFERÊNCIAS GERAIS	92

1. INTRODUÇÃO GERAL

O Brasil é referencia na silvicultura do *Eucalyptus* e esse mérito é devido aos grandes avanços no melhoramento genético tradicional, bem como em melhores práticas de manejo silviculturais desenvolvidas. A área plantada com *Eucalyptus* no país em 2015 foi de 5,56 milhões de hectares sendo que 11,6% da área estão concentrados na região Sul do país (IBÁ, 2016).

O *Eucalyptus benthamii* Maiden et Cambage é uma espécie de grande importância devido a sua alta tolerância ao frio aliada à boa adaptação e crescimento volumétrico. Porém, o melhoramento genético dessa espécie é recente e a principal estratégia usada atualmente tem sido a Seleção Recorrente Intrapopulacional (SRI) para o melhoramento da espécie, visto que essa espécie é referencia em termos de tolerância a geada em regiões com variações drásticas de temperatura. Outra estratégia adotada é a Seleção Recorrente Recíproca em População Sintética Multi-espécie (SRR-PSME). Polinizações controladas do *E. benthamii* com outras espécies ou clones superiores também tem sido realizada com a finalidade de agregar bom crescimento e qualidade da madeira à tolerância a geada dessa espécie (FONSECA et al., 2010).

A seleção de indivíduos superiores em programas de melhoramento genético vem sendo praticada com base no valor genotípico estimado a partir de avaliações fenotípicas em poucas características mensuradas em campo. Para aumentar os ganhos genéticos nos programas de melhoramento foi proposta e desenvolvida na década de 1990 a seleção assistida por marcadores moleculares (Marker assisted selection – MAS). Foram realizados trabalhos para diversas culturas, inclusive para o *Eucalyptus*, estudos de QTL visando crescimento, resistência a doença (JUNGHANS et al., 2003), qualidade da madeira (THAMARUS et al., 2004) ou florescimento precoce (MISSIAGIA et al., 2005), houve um progresso considerável com esta técnica, porém a

implementação foi limitada, revisões sobre o tema foram citadas por GRATTAPAGLIA (2014).

Um dos problemas dessa técnica foi que um marcador para ser útil precisa explicar grande parte da variação genética de uma característica quantitativa, as quais são controladas por vários genes de pequenos efeitos e, geralmente é encontrado um pequeno número de QTLs (“Quantitative Trait Loci”) de grandes efeitos, que não explicam toda a variação genética de forma efetiva (DEKKERS, 2004; BERNARDO, 2008).

Como alternativa à seleção assistida por marcadores foi proposta em 2001 a Seleção Genômica (SG), com o objetivo de acelerar o ciclo de seleção nos programas de melhoramento genético, com potencialidade de aumentar a acurácia seletiva.

Com o desenvolvimento da SG pode ser possível eliminar algumas etapas dos programas de melhoramento, como o teste de progênie e após consistências e validações práticas até mesmo o teste clonal. Os futuros clones poderão ser recomendados com um tempo mínimo, somente para a avaliação em campo quanto à competição intraespecífica, nos denominados plantios pilotos ou testes clonais ampliados. Essa inovação tecnológica será muito importante para o setor florestal, pois irá acelerar e otimizar os ganhos de seleção, aumentando ainda mais a competitividade do Brasil em floresta plantada de *Eucalyptus* (GRATTAPAGLIA, 2014).

O presente trabalho teve como objetivo geral validar tecnologias que visem acelerar o melhoramento genético do *E. benthamii*, para tal em dois capítulos apresentaram-se as seguintes realizações: i) Predição genômica para volume e qualidade da madeira para o melhoramento de *E. benthamii*. ii) Caracterização não destrutiva da madeira de *E. benthamii* por meio de espectroscopia do infravermelho próximo (NIR).

2. REVISÃO DE LITERATURA

2.1 *EUCALYPTUS BENTHAMII* E O MELHORAMENTO GENÉTICO

O *Eucalyptus benthamii* é uma espécie de grande importância no sul do Brasil devido a sua alta tolerância ao frio aliada à boa adaptação e crescimento volumétrico (COSTA et al., 2016). Essa espécie tem sua ocorrência natural na Austrália no litoral de New South Wales, a sudoeste da cidade de Sydney, nas planícies ao longo do Rio Nepean e seus afluentes. É considerada uma espécie vulnerável ao desaparecimento na região de ocorrência natural. Restam apenas quatro populações, sendo a maior delas constituída por 6.550 árvores e as outras três com menos de 340 árvores. As sementes dessas populações têm pouca viabilidade face ao alto grau de autofecundação e endogamia. Atualmente essas áreas de ocorrência foram transformadas em parques nacionais, onde não são permitidas coletas do material genético, o que reforça o cuidado no melhoramento e na conservação da base genética da espécie existente no Brasil (BUTCHER et al., 2005).

O melhoramento do *Eucalyptus* no Brasil é praticado principalmente por indústrias de celulose e indústrias siderúrgicas. Outras instituições que contribuem para o melhoramento do *Eucalyptus* são as universidades UFV, ESALQ/USP, UFLA, UFPR e UNESP, as quais geram teses e estudos referentes a esse tema (FONSECA et al., 2010).

Atualmente, o melhoramento de *Eucalyptus* se baseia na seleção de genitores e no direcionamento dos cruzamentos dentro de programas de SRR entre populações divergentes e também via SRI em população híbrida sintética oriunda do cruzamento entre clones elites com incorporação de novos clones a cada geração. Esse último procedimento tem sido denominado Seleção Recorrente Intrapopulacional em

População Sintética (SRIPS), população essa “re-sintetizada” a cada geração via cruzamentos controlados (RESENDE & BARBOSA, 2005).

Com a expansão das fronteiras agrícolas e o plantio em áreas com certas adversidades, surgiu a necessidade de inclusão de outras espécies, além de *E. grandis* e *E. urophylla*, nos cruzamentos visando a combinação em um único indivíduo, das características: alto crescimento, alta densidade da madeira, alto rendimento de celulose, tolerância à seca, resistência à doenças e, em alguns casos, tolerância ao frio. Essas espécies são *E. globulus*, *E. camaldulensis*, *E. pellita*, *E. dunnii*, *E. benthamii*, *E. viminalis*, essa nova estratégia de melhoramento SRR-PSME é denominada Seleção Recorrente Recíproca em Populações Sintéticas Multi-Espécies (FONSECA et al., 2010).

O programa de melhoramento genético do *E. benthamii* é recente, as estratégias mais usadas são a SRI para o melhoramento da espécie pura, visto que atualmente essa espécie é referência em termos de tolerância a frio. Polinizações controlados de *E. benthamii* com outras espécies ou clones superiores também tem sido realizadas com a finalidade de agregar a tolerância à geada aos híbridos (RESENDE & ASSIS, 2008; FONSECA et al., 2010).

2.2 SELEÇÃO GENÔMICA E A SELEÇÃO ASSISTIDA POR MARCADORES

A seleção de indivíduos superiores em programas de melhoramento genético vem sendo praticada a partir de avaliações fenotípicas de poucas características mensuradas em campo. Para otimizar os ganhos genéticos nos programas foi proposto por Lande e Thompson (1990), a Seleção Assistida por Marcadores (SAM). Muitos trabalhos utilizando a seleção assistida por marcadores foram desenvolvidos e o mapeamento de QTLs foram realizados para diversas culturas, inclusive arbóreas

(JUNGHANS et al., 2003; THAMARUS et al., 2004; MISSIAGIA et al., 2005)) porém a implementação da SAM tem sido limitada e os ganhos em eficiência muito reduzidos com resultados ineficientes para caracteres quantitativos, devido as seguintes características da SAM: (i) requer o estabelecimento de associações marcadores-QTLs para cada família em estudo, ou seja, essas associações apresentam utilidade para seleção apenas dentro de cada família mapeada em espécies alógamas; (ii) para ser útil precisa explicar grande parte da variação genética de uma característica quantitativa, que em geral é controlada por muitos locos de pequenos efeitos. Em função da natureza poligênica e alta influência ambiental nos caracteres quantitativos, são detectados apenas um pequeno número de QTLs de grandes efeitos, os quais não explicam suficientemente toda a variação genética; (iii) somente apresenta superioridade em relação à seleção baseada em dados fenotípicos, quando o tamanho de família avaliada e genotipada é muito grande, da ordem de 500 ou mais (DEKKERS, 2004; BERNARDO, 2008).

Com objetivo de aumentar a acurácia seletiva e acelerar os ganhos genéticos nos programas de melhoramento, foi proposta a seleção genômica por MEUWISSEN et al. (2001), porém a técnica ficou em repouso por alguns anos em função dos altos custos de genotipagem. A seleção genômica foi denominada ampla por atuar em todo o genoma sem a necessidade prévia de identificar marcadores moleculares com efeitos significativos e de mapear QTLs, podendo ser aplicada em todas as famílias em avaliação nos programas de melhoramento genético de espécies alógamas, apresentando alta acurácia seletiva para a seleção baseada exclusivamente em marcadores e não exigindo o prévio conhecimento das posições (mapa) dos QTLs, não estando sujeita aos erros tipo I de aceitar uma hipótese que é falsa (RESENDE, 2007a).

A SAM baseia-se na detecção, mapeamento e no uso de QTLs de grande efeito na seleção, após determinar o número, a posição e os efeitos dos QTLs marcados (LANDE & THOMPSON, 1990). Enquanto a seleção genômica é definida como a seleção simultânea para centenas ou milhares de marcadores, que cobrem o genoma de uma maneira densa, de forma que todos os genes de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores. Esses marcadores em desequilíbrio de ligação com os QTLs, tanto de grandes quanto de pequenos efeitos, explicarão quase a totalidade da variação genética de um caráter quantitativo. O número de marcadores usado deve ser de tal magnitude que a probabilidade de se encontrar um QTL em desequilíbrio de ligação com pelo menos um marcador seja muito alta. Esse aspecto é importante uma vez que somente os marcadores em desequilíbrio de ligação com os QTLs são úteis na determinação dos fenótipos e na explicação da variação genética (RESENDE, 2007a; BERNARDO, 2008).

Os efeitos dos marcadores são estimados em uma amostra de indivíduos pertencentes a várias famílias. Assim, o impacto de determinadas famílias específicas (com específicos padrões de desequilíbrio de ligação) nas estimativas dos efeitos dos marcadores será minimizado. É importante enfatizar que os marcadores terão seus efeitos genéticos estimados a partir de uma grande amostra de indivíduos genotipados e fenotipados, ou seja, com base em várias repetições experimentais. Assim, embora a herdabilidade de cada marcador efetivo seja muito baixa, com muitas repetições essa herdabilidade se torna elevada (RESENDE et al., 2008).

Em outras palavras, o efeito de ambiente será minimizado por meio do uso de um número de repetições muito alto. Essa é a mesma filosofia da avaliação e seleção de características quantitativas com base em fenótipos em experimentos de campo (RESENDE et al., 2008).

Um estudo sobre o impacto da SG no melhoramento do *Eucalyptus* utilizando marcadores dominantes DArTs e por meio de simulações e de dados reais, demonstrou o grande potencial da técnica se comparados a magnitude das acurácias dos ganhos e da diminuição do tempo de seleção (RESENDE JÚNIOR, 2010). Diversos outros trabalhos mostraram que SG pode ser uma estratégia muito eficiente para os programas de melhoramento genético (BERNARDO & YU, 2007; MEUWISSEN, 2007; DAETWYLER et al., 2008, CROSSA, 2010; IWATA, 2011; GRATTAPAGLIA & RESENDE, 2011; DE LOS CAMPOS et al., 2013).

Vários estudos têm analisado as perspectivas da aplicação da SG em animais, considerando parâmetros e procedimentos analíticos que influenciam na predição dos valores genômicos, além do impacto de seu uso no melhoramento (SCHAEFFER 2006; DEKKERS, 2007; LONG et al., 2007; MUIR, 2007; CALUS et al., 2008; SOLBERG et al., 2008). Para algumas espécies animais, como exemplo gado leiteiro, a SG é uma realidade prática mostrando resultados de altas acurácias seletivas e a diminuição no tempo de melhoramento, sendo possível redução de tempo do teste de progênie (PRYCE et al., 2012).

Em plantas, essa técnica também possui grande potencial, visto que, pesquisas recentes, com dados simulados e reais, demonstraram excelentes resultados na utilização dessa tecnologia em diferentes culturas (BERNARDO & YU 2007; RESENDE et al., 2008; WONG & BERNARDO 2008; GORE et al., 2009; HEFFNER et al., 2009; ZHONG et al., 2009). Em espécies florestais, os primeiros resultados obtidos no melhoramento são bastante promissores, porém ainda necessitam de mais estudos para se tornar uma realidade nos programas de melhoramento (RESENDE et al., 2012; LIMA, 2014; REZENDE, 2015; GARCIA, 2016).

A SG propicia uma forma de seleção precoce direta, atuando sobre genes que

muitas vezes são expressos somente em idade adulta, enquanto a seleção precoce tradicional é indireta, pois, atua sobre genes ativados na idade precoce por meio da avaliação dos fenótipos, esperando que esses informem parcialmente sobre genes expressos na idade adulta. A SG é especialmente importante para acelerar o melhoramento de espécies perenes como animais, espécies florestais, frutíferas, forrageiras, cana-de-açúcar, dentre outras, onde a mensuração é tardia para algumas características (RESENDE, 2008).

Essa tecnologia é superior em relação à seleção fenotípica realizada atualmente devido a quatro fatores: (i) uso da matriz de parentesco real e própria de cada caráter, fato que aumenta a acurácia seletiva; (ii) viabilização da seleção precoce, que aumenta o ganho genético por unidade de tempo; (iii) permite avaliação repetida de cada alelo, com alta repetição experimental, fato que aumenta a acurácia seletiva; (iv) uso de maior número de informações, combinando três tipos de informação (fenotípica, genotípica e genealógica) para corrigir e desregressar os dados e fazer a análise genômica, fato que aumenta a acurácia (RESENDE, 2007a). Há uma diferença básica na predição de valores genotípicos e na predição de valores genéticos genômicos (VGG). Nos primeiros, informações fenotípicas são utilizadas visando inferências sobre os efeitos dos genótipos dos indivíduos e, nos últimos, as informações genotípicas (genótipos para os alelos marcadores) são usadas visando às inferências sobre os valores fenotípicos futuros (ou valores genéticos genômicos preditos) dos indivíduos. Em outras palavras, os métodos tradicionais usam o fenótipo para inferir sobre o efeito do genótipo e a SG usa o genótipo, com efeito genético pré-estimado em uma amostra da população, para inferir sobre o fenótipo a ser expresso nos candidatos à seleção (RESENDE et al., 2008).

Os efeitos dos marcadores não serão necessariamente os mesmos em diferentes estudos e ambientes. Na SG, os efeitos genéticos dos marcadores são estimados e usados na seleção para cada população de melhoramento e em um determinado ambiente. E os modelos de estimação, incluindo a interação genótipos x ambientes, podem também ser usados, visando verificar a possibilidade de se obter estimativas válidas para um conjunto de ambientes. Porém, isso dependerá da magnitude da interação envolvendo os vários ambientes (RESENDE et al., 2008).

A SG pode basear-se no uso de: (i) apenas dos marcadores; (ii) de haplótipos ou intervalos definidos por dois marcadores; (iii) haplótipos definidos por mais de dois marcadores, incluindo a covariância entre haplótipos devida à ligação. Porém, para caracteres de baixa herdabilidade (10%) não existem diferenças significativas entre essas três abordagens (CALUS et al., 2008).

Na SG não apenas os marcadores SNPs podem ser usados, mas também os microssatélites. O uso de SNPs requer quatro a cinco vezes maior densidade de marcadores do que o uso de microssatélites. Isso se deve à natureza bi-alélica (binucleotídica) dos SNPs e multi-alélica dos microssatélites (SOLBERG et al., 2008). Marcadores dominantes também podem ser usados na SG e atingem acurácias satisfatórias, a exemplo do DArt (RESENDE et al., 2012; REZENDE, 2015).

2.3 O PROCEDIMENTO RR-BLUP GENÔMICO

Para estimar os valores genéticos genômicos dos indivíduos é usado um conjunto de dados de referência que inclui: genótipos (marcadores moleculares) e os fenótipos, denominada população de estimação. Os valores genéticos estimados dos haplótipos ou marcadores por um grande número de caracteres quantitativos são usados para a predição dos valores genéticos genômicos de indivíduos jovens candidatos à

seleção e que foram genotipados para os marcadores, mas não possuem informação fenotípica. Se toda a variação genética puder ser explicada pelos haplótipos ou marcadores, não há necessidade de inclusão no modelo de predição, do efeito poligênico para levar em consideração a variação genética não explicada (variação genética residual) (RESENDE, 2007a).

Na prática, se não há uma cobertura completa (mapa denso de marcadores) do genoma com SNPs, a inclusão do efeito poligênico pode tornar-se necessária. A estimação dos valores genéticos genômicos para haplótipos marcadores individuais ou alelos individuais do QTL baseia-se em um número relativamente grande de haplótipos e outro relativamente pequeno de indivíduos (RESENDE, 2007a).

O problema da metodologia dos quadrados mínimos na SG é que, o número de marcas cujos efeitos precisam ser estimados é muitas vezes maior que o número de observações fenotípicas, não existem graus de liberdade suficientes para ajustar todos estes efeitos simultaneamente, sendo necessário estimar um efeito de cada vez e verificar sua significância. Além disso, os marcadores são altamente correlacionados o que conduz a problemas de multicolinearidade. O método RRBLUP, por outro lado, permite ajustar todos os efeitos dos marcadores simultaneamente, mesmo quando existem mais efeitos a serem preditos do que o número total de observações fenotípicas, uma vez que considera os efeitos dos marcadores como efeitos aleatórios (RESENDE et al., 2008).

Os efeitos estimados dos marcadores são assumidos como estimativas válidas para toda a população e não para apenas um grupo de indivíduos. Dessa forma, VGG podem ser estimados para quaisquer indivíduos da população, desde que os mesmos sejam genotipados e os marcadores sejam determinados. Assim, cada indivíduo pode ter uma estimativa do VGG desde o momento em que é gerado. Os efeitos de cada

marcador em um caráter em um dado ambiente podem ser estimados para todos os marcadores simultaneamente em um modelo linear misto em que os efeitos de intervalo são tratados como aleatórios. Os genótipos marcadores dos indivíduos podem ser usados para predição de qualquer caráter, mas as estimativas dos efeitos dos intervalos ou haplótipos serão diferentes para cada caráter. Os intervalos com maiores efeitos em cada caráter conterão um ou mais QTLs. A maioria dos intervalos, no entanto, apresentará efeitos relativamente menores, refletindo o que acontece no modelo infinitesimal (muitos genes de pequenos efeitos associados ao caráter quantitativo).

Para o uso dos SNPs na SG, inicialmente, devem ser identificados àqueles informativos e, posteriormente, um software deve ser usado para construir haplótipos a partir dos genótipos SNP. De posse dos haplótipos, as predições de seus efeitos podem ser feitas por meio de softwares específicos de genética quantitativa e estatística (RESENDE, 2007b).

O seguinte modelo linear misto geral é usado para estimar os efeitos de marcadores:

$$y = Xb + Zm + e$$

$$m \sim N(0, I\sigma^2_m)$$

em que: y é o vetor de observações fenotípicas, b é o vetor de efeitos fixos, m é o vetor dos efeitos aleatórios de marcadores e refere-se ao vetor de erros aleatórios.

X e Z são as matrizes de incidência para b e h .

A matriz de incidência Z contém os valores 0, 1 e 2 para o número de alelos (do SNP) ou marcadores do tipo h_i no indivíduo diplóide j .

As equações genômicas de modelo misto para a predição de m via o método RR-BLUP equivalem a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que σ_g^2 refere-se à variância genética total do caráter σ_e^2 é a variância residual e n é o número de QTLs que pode ser estimado via $n_Q = \sum_{j=1}^n 2p_j(1-p_j)$, em n_Q é o número de locos marcadores, em que p_j é a frequência de um marcador o loco j .

O valor genético genômico global do indivíduo j é dado por:

$$VGG = \hat{y}_j = \sum_i Z_i \hat{m}_i$$

Em que o componente Z_{ij} refere-se ao elemento i da linha j da matriz Z , referente ao indivíduo, em que Z_i equivale a 0, 1 ou 2.

As equações de predição apresentadas anteriormente assumem, *a priori*, que todos os locos explicam iguais quantidades da variação genética. Assim, a variação genética explicada por cada loco é dada por (σ_g^2/n) , em que σ_g^2 é a variação genética total e n é o número de marcadores. A variação genotípica σ_g^2 pode ser estimada por REML sobre os dados fenotípicos da maneira tradicional ou pela própria variação entre marcadores ou dos segmentos cromossômicos de QTL.

No procedimento do RR-BLUP não há necessidade de uso da matriz de parentesco, baseada em pedigree. Essa é substituída por uma matriz de parentesco estimada pelos marcadores (RESENDE, 2007a).

2.4 IMPLEMENTAÇÃO DA SELEÇÃO GENÔMICA

Na prática a seleção genômica necessita de três conjuntos de dados conforme descrito na sequência, com base em GODDARD & HAYES (2007). População de Descoberta: Esse conjunto de dados contempla grande número de marcadores SNPs avaliados em um número moderado de indivíduos, os quais devem ter seus fenótipos avaliados com relação a vários caracteres de interesse. Equações de predição de valores genéticos genômicos são obtidas para cada caráter de interesse. Essas equações associam a cada intervalo marcador o seu no caráter de interesse. Nessa população são descobertos, via marcadores, os locos que controlam os caracteres, bem como são estimados os seus efeitos.

População de Validação: Esse conjunto de dados é menor que aquele da população de descoberta e contempla indivíduos avaliados para os marcadores SNPs e para os vários caracteres de interesse. As equações de predição de valores genéticos genômicos são testadas para verificar suas acurácias nessa amostra independente. Para computar essa acurácia, os valores genéticos genômicos são preditos (usando-se os efeitos estimados na população de descoberta) e submetidos à análise de correlação com os valores fenotípicos observados. Como a amostra de validação não foi envolvida na predição dos efeitos dos marcadores, os erros dos valores genéticos genômicos e dos valores fenotípicos são independentes, e toda correlação entre esses valores é de natureza genética e equivale à própria acurácia. Cerca de 200 indivíduos híbridos de uma população de melhoramento elite são suficientes para compor a população de validação. Esses devem ser diferentes daqueles empregados na população de descoberta.

População de Seleção: Esse conjunto de dados contempla apenas os marcadores SNPs avaliados em todos os candidatos à seleção. Essa população não necessita ter os seus fenótipos avaliados. As equações de predição derivadas na população de

descoberta são, então, usadas na predição dos VGGs, ou fenótipos futuros dos candidatos à seleção. Mas a acurácia seletiva associada se refere àquela calculada na população de validação.

De acordo com MEUWISSEN (2007), quando dezenas e centenas de milhares de marcadores são estimados, há o risco de superparametrização, ou seja, de erros nos dados serem explicados pelos efeitos de marcadores. A validação cruzada é, então, de grande importância para contornar esse problema.

Sob seleção genômica, todos os candidatos à seleção (indivíduos sem observação fenotípica) poderão ser avaliados para quaisquer ambientes, desde que tais ambientes possuam equações de predição derivadas para os próprios e com alta acurácia. Assim, indivíduos poderão ser selecionados com base em seus valores fenotípicos preditos, derivados de um catálogo de marcadores associados aos candidatos à seleção.

Para a implementação da SG, uma população de melhoramento com tamanho efetivo entre 30 a 100 suporta a seleção com ganho genético considerável por várias gerações. Esse poderá conduzir a aumento da ordem de 100% na eficiência seletiva, com redução de cerca de 50% no intervalo entre gerações de melhoramento, ou seja, no ciclo seletivo. Assim, pode ser aplicada vantajosamente em qualquer geração de melhoramento (FONSECA et al., 2010).

2.5 ESTIMADOR DA ACURÁCIA DE SELEÇÃO GENÔMICA

A acurácia é a principal medida para avaliar a eficiência dos valores genéticos genômicos. Um estimador tradicional para a precisão experimental do GWS foi introduzido por LEGARRA et al. (2008) e HAYES et al. (2009). Tal estimador é dado pela relação entre a capacidade de predição e a raiz quadrada da herdabilidade da

característica. No entanto, em algumas circunstâncias, este estimador tem inconsistências como o fato de uma população sem variabilidade genética (herdabilidade da característica igual a zero) conduzir a uma acurácia elevada na GWS e além disso, as estimativas da acurácia podem estar fora do espaço paramétrico (maior que 1).

Desta forma, outro estimador de acurácia foi descrito por RESENDE et al. (2014), também apresentado por VIANA & RESENDE (2014), e validado por Azevedo et al. (2016) uma alternativa para estimar a acurácia da seleção genômica de forma mais consistente por estar dentro do espaço paramétrico. Esse estimador proposto é dado por $rgg^* = (rgg) / \sqrt{(h^2_t)} \cdot \sqrt{(h^2_M)}$, ou seja, $rgg^* = rgf \times hM / h$ ou $rgg^* = rgg \times hM$, em que hM é a raiz quadrada da herdabilidade genômica ou molecular e x é o operador de multiplicação. Esses estimadores multiplicam a proximidade (distância) entre g e f (dada por rgf), pela proximidade entre g e gv (dada por hM/h), produzindo o que se deseja, ou seja, a proximidade entre g e gv , o valor genético verdadeiro. Nesse caso, multiplicar rgf por hM/h diminui a acurácia, ou seja, penaliza rgf , como deve ser. Adicionalmente $rgg^* = rgf \times hM / h$ possui propriedades estatísticas adequadas, pois, produz rgg^* no espaço paramétrico de 0 a 1. Em $rgg^* = rgf \times hM / h$, quanto menor hM/h , menor rgg^* , fato que é condizente com a realidade. Assim, as inconsistências da fórmula tradicional são corrigidas. Esse estimador é conservador (produz menores acurácias que a fórmula tradicional) e é função de três parâmetros, usando também a herdabilidade genômica. Essa hM como proporção de h , leva em conta a eficiência dos marcadores em capturar os QTL, ou seja, considera o grau de imperfeição do desequilíbrio de ligação.

2.6 A TECNOLOGIA DE ESPECTROSCOPIA DO INFRAVERMELHO PRÓXIMO (NIR)

A amostragem não destrutiva da madeira é definida como sendo a ciência de identificar as propriedades físicas e mecânicas de uma peça de determinado material sem alterar suas capacidades de uso final (ROSS et al., 1998). Esse tipo de análise é de grande importância para os programas de melhoramento genético, visto que muitas vezes não é possível abater a árvore que é um material genético único e de valor para o avanço das gerações. Várias metodologias de amostragem não destrutivas são aplicadas atualmente com o objetivo de prever as propriedades químicas e físicas da madeira por meio de modelos de predição. A espectroscopia NIR tem se destacado nas indústrias de produção de celulose devido ao seu suporte às pesquisas de caracterização do produto e processo (MILAGRES et al., 2013).

De maneira geral, a espectroscopia é o estudo da interação da radiação eletromagnética com a matéria. A radiação eletromagnética se distribui em diversas regiões de energia que correspondem a diferentes técnicas de espectroscopia. As principais técnicas de espectroscopia são de infravermelho, visível e ultravioleta. A região espectral do infravermelho próximo compreende a faixa de 4.000 a 10.000 cm^{-1} , sendo que as amostras apresentam absorvência muito menor que na região tradicional do infravermelho médio (EISBERG & RESNICK, 1986). A espectroscopia no infravermelho próximo é a medição da intensidade de luz infravermelha em relação ao comprimento de onda emitido pela amostra. É tipicamente usada na medição quantitativa de absorção/reflexão e pode ser determinado com um espectrofotômetro, que consiste em uma fonte luminosa, um monocromador, que contém o seletor de comprimentos de ondas dos grupos funcionais orgânicos, especialmente O-H, N-H, e

C=O, conforme apresentado na Figura 1. O gráfico de reflectância ($\log 1/R$) *versus* comprimento de onda (λ) é denominado espectro. E a partir desses espectros, juntamente com os dados de referência, são construídos os modelos para predição (NISGOSHI, 2005).

O NIR é uma tecnologia que vem sendo usada para diversas finalidades em pesquisas de laboratórios e nos processos industriais, é um método rápido e barato, que tem contribuído bastante para as determinações químicas (RAYMOND; SOUSA, 2008; VIANA et al., 2010; MILAGRES, 2013, RAMADEVI et al., 2016) e até mesmo físicas de componentes importantes da madeira (SCHIMLECK et al., 2005; HEIN et al., 2009, ROSSO et al., 2013).

Fantuzzi Neto, 2012 desenvolvendo modelos de predição para eucalipto aos 7 anos alcançou coeficientes de determinação do modelo de 0,87, 0,95 e 0,90 para densidade básica da madeira, lignina e extrativos, respectivamente. Tais modelos apresentaram boas capacidades preditivas, sendo R^2 foi 68% e erro de 0,016 g/cm³ para densidade básica. Para o modelo de lignina o coeficiente de determinação da predição foi de 0,71 e erro de 1,09%, o modelo de extrativos apresentou R^2 de 0,84 e erro de 0,5%.

O único trabalho utilizando NIRS para predizer a densidade básica da madeira de *E. benthamii* encontrado na literatura foi realizado por Lazzarotto & Magalhães (2014), para calibrar o modelo de densidade básica da madeira foi utilizado 40 árvores de *E. benthamii* e 44 de *E. pellita* e, obtiveram boas correlações, maior que 0,8, para a calibração. O erro padrão de predição encontrado foi de 0,027 g/cm³, coeficiente de determinação de predição de 0,62. Pasquini et al. 2007, trabalhando com 926 amostras de eucalipto para calibração e 126 para validação externa obteve erro de predição semelhante ao encontrado neste trabalho 0,026 g/cm³ e coeficiente de determinação de

0,86.

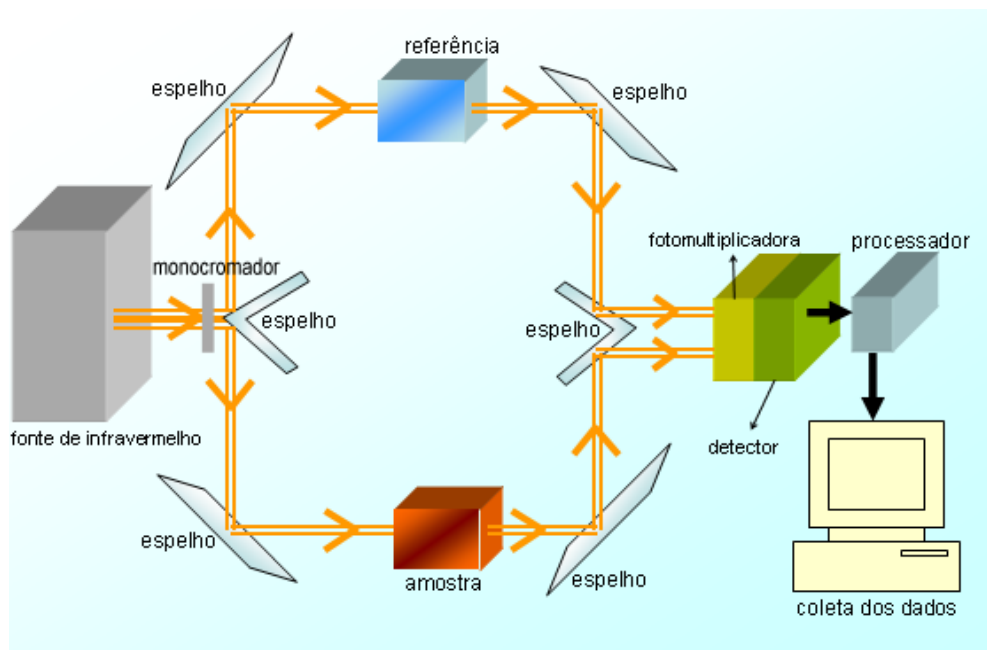


Figura 1 – Esquema do princípio operacional de um sistema de espectrofotometria com transformada de Fourier*.

*Disponível em: (<<https://pt.wikipedia.org/wiki/espectroscopiadeinfravermelho>> acesso em fev. de 2017).

Considerando a necessidade seleção precoce e não destrutiva, a espectroscopia de infravermelho próximo (NIR) vem sendo usada para diversos fins em pesquisa e processos industriais. Tem sido apontada como um método rápido e barato (RAYMOND & SCHIMLECK, 2002, SOUSA, 2008, VIANA et al., 2010, MILAGRES, 2013, RAMADEVI et al., 2016), para análise física dos principais componentes da madeira (SCHIMLECK et al., 2005; HEIN et al. 2009; ROSSO et al., 2013), produtividade de polpa (SANTOS et al., 2009) e até propriedades de papel (SAMISTRARO et al., 2009). Outra aplicação de NIR em árvores foi indicar que a tecnologia pode ser usada como uma ferramenta eficaz para distinguir entre espécies puras de pinus e sugerem que também poderá distinguir híbridos desses parentais (ESPINOZA et al., 2012). Já OLIVEIRA et al. (2015), por meio de espectroscopia NIR e redes neurais artificiais conseguiu distinguir quatro espécies arbóreas. Também foi

possível distinguir espécies arbóreas nativas destinadas à produção de carvão (DAVRIEUX et al., 2010, NISGOSKI et al., 2015).

Em relação à seleção de árvores em programas de melhoramento baseado em informações de espectroscopia NIR, muitos estudos apresentaram resultados promissores. GREAVES et al. (1996) avaliaram o controle genético de espectros NIR registrados em serragem de madeira de 588 indivíduos de *Eucalyptus nitens*, enquanto HEIN e CHAIX (2014) estudaram espectros NIR em madeira sólida de clones híbridos *Eucalyptus urophylla* x *E grandis*. Em ambos os estudos, calcularam a razão entre a variância genética e a variância fenotípica dos valores de absorvência NIR para cada número de onda que produz estimativas de herdabilidade no sentido amplo para espectros NIR. Eles relataram que alguns intervalos dos espectros capturados na madeira de clones de Eucalyptus aos 6 anos apresentaram estimativas de herdabilidade superiores a 0,50, indicando que o melhorista poderia selecionar árvores potenciais utilizando os espectros NIR mais herdáveis (HEIN, CHAIX, 2014).

3. REFERÊNCIAS

AZEVEDO, C.F.; RESENDE, M.D.V.; SILVA, F.F.; VIANA, J.M.S.; VALENTE, M.S.F.; RESENDE JR, M.F.R.; OLIVEIRA, E.J. New accuracy estimators for genomic selection with application in a cassava (*Manihot esculenta*) breeding program. **Genetics and Molecular Research**, v. 15, p. 1, 2016.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, 2008.

BERNARDO, R; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, n.3, p.1082-1090, 2007.

BUTCHER, P.A., SKINNER, A.K., GARDINER, C.A. Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii*. **Conservation Genetics**, v. 6, n. 2, p.213–226, 2005.

COSTA, R.M.L.; ESTOPA, R.A.; BIERNASKI, F.A.; MORI, E.S. Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cambage por diferentes métodos de seleção. **Scientia Forestalis**, v. 44, n. 109, p. 105-113, 2016.

CROSSA J.; DE LOS CAMPOS, G.; PEREZ, P. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v.186, n.2, p.713-724, 2010.

DAVRIEUX, F.; ROUSSET, P.L.A.; PASTORE, M.C.T.; MACEDO, L.A. and QUIRINO, W.F. Discrimination of native wood charcoal by infrared spectroscopy. *Química Nova*, v. 33, p. 1093-1097, 2010.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, v.193, p.327-345, 2013.

DEKKERS, J.C.M. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, n.13, p.313-328, 2004.

DEKKERS, J.C.M. Marker-assisted selection for commercial crossbred performance. **Journal of Animal Science**, v. 85. p.2104-2114, 2007.

EISBERG, R.; RESNICK, R. **Física quântica**: átomos, moléculas, sólidos, núcleos e partículas. 4º ed. Rio de Janeiro: Campus, 372p, 1986.

ESPINOZA, J. A.; HODGE, G.R.; DVORAK W.S. The potential use of Near Infrared Spectroscopy to discriminate between pine spices and their hybrids. *Journal of Near Infrared Spectroscopy*, v. 18, p. 437-447, 2012.

FONSECA, S.M.; RESENDE, M. D. V.; ALFENAS, A. C.; GUIMARAES, L. M. S.; ASSIS, T. F. ; GRATTAPAGLIA, D. **Manual prático de melhoramento genético do *Eucalyptus***. Viçosa: Editora UFV, 200p, 2010.

GARCIA, C.C. Avaliação de modelos preditivos de seleção genômica ampla em testes de progênies e testes clonais de *Eucalyptus*. 2016, 92p. **Dissertação** (Mestrado em Ciências: Recursos Florestais) – Escola Superior de Agricultura/USP, Piracicaba.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 323-330, 2007.

GORE, M.A.; CHIA J.M.; ELSHIRE, R.J. et al. A First-Generation Haplotype Map of Maize. **Science**, v. 326, n. 5956, p. 1115-1117, 2009.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. (Ed.). **Genomics of plant genetic resources**. Dordrecht: Springer, 2014. chap. 26, p. 651-682.

GRATTAPAGLIA, D.; RESENDE M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.

GREAVES; B.L.; SCHIMLECK, L.R.; BORRALHO, N.M.G. and MICHELL, A.J. Genetic control and repeatability of near infrared reflectance from *Eucalyptus nitens* woodmeal, **Appita J.** v. 49, p. 423-426, 1996.

HAYES, B.J.; VISSCHER, P.M.; GODDARD, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, v. 91, n. 1, p. 47-60, 2009.

HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J.L. Genomic selection for crop improvement. **Crop Science**, v. 49, n. 1, p.1–12, 2009.

HEIN, P.R.G. and CHAIX, G. NIR spectral heritability: a promising tool for wood breeders. **Journal of Near Infrared Spectroscopy**, v. 22, p. 141-147, 2014.

HEIN, P.R.G.; LIMA, J.T.; Chaix, G. “Robustness of models based on near infrared spectra to predict the basic density in *Eucalyptus urophylla* wood”. **Journal of Near Infrared Spectroscopy**, v.17, p.141-150, 2009.

IBÁ. **Indústria Brasileira de Árvores**, ano base 2015, Brasília, 2016.

IWATA, H.; HAYASHI, T.; TSUMURA, Y. Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. **Tree Genetics & Genomes**, v.7, n.4, p. 747-758, 2011.

JUNGHANS, D.T.; ALFENAS, A.C.; BROMMONSCHENKEL, S.H.; ODA, S.; MELLO, E.J.; GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of major gene with RAPD markers. **Theoretical and Applied Genetics**, v. 108, n. 1, p. 175-180, 2003.

LANDE, R.; THOMPSON, R. Efficiency of marker assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n. 3, p. 743-756, 1990.

LEGARRA A., ROBERT-GRANIE C., MANFREDI E., ELSSEN J. M. Performance of genomic selection in mice. **Genetics**, v. 180, p. 611-618, 2008.

LIMA, B.M. Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data. 2014, 92p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura/USP, Piracicaba.

LONG, N.; GIANOLA, D.; ROSA, G. J.M.; WEIGEL, K.A; AVENDAÑO, S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, v. 124, n.6, p. 377-389, 2007.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p.1819-1829, 2001.

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 321-322, 2007.

MILAGRES F. R.; GOMIDE, J. L.; MAGATON, A. S.; FANTUZZI NETO H. The influence of the age on near infrared spectroscopy models to prediction of the *Eucalyptus spp.* wood properties. **Revista Árvore**, v. 37, p.1165-1173, 2013.

MISSIAGIA, A. A.; PIACEZZI, A. L.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics and Genomes**, Heidelberg, v. 1, n. 1, p. 79-84, Oct. 2005.

MUIR, W.M. Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, v. 124, n.6, p. 342-355, Dec. 2007.

NISGOSKI, S.; MUÑIZ, G.I.B.; MORRONE, S.R.; SCHARDOSIN, F.Z.; FRANÇA, R.F. NIR and anatomy of wood and charcoal from Moraceae and Euphorbiaceae species. *Ciência da Madeira* v. 6, n. 3, p. 183-190, 2015.

PRYCE, J.E.; DAETWYLER, H.D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. **Animal Production Science**, v.52, n. 3, p.107-114, Dec. 2012.

RAMADEVI, P.; HEGDE, D.V.; VARGHESE, M.; KAMALAKANNAN, R, GANAPA-THY, S.P. and GURUMURTHY, D.S. Evaluation of lignin syringyl/guaiacyl ratio in *Eucalyptus camaldulensis* across three diverse sites based on near infrared spec-troscopic calibration modelling with five *Eucalyptus* species and its impact on Kraft pulp yield. **J. Near Infrared Spectroscopy**, v. 24, p. 529-536, 2016.

RAYMOND, C.A. and SCHIMLECK, L.R. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. *Canadian Journal of Forest Research*, v. 32, p. 170–176, 2002.

RESENDE, M.D.V. DE; SILVA, F. F. E. ; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência**. 1. ed. Visconde do Rio Branco: Suprema, 2014, v. 1. 881p.

RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RESENDE, M.D.V. Selegen-REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Colombo: **Embrapa Florestas**, 2007b. 359p.

RESENDE, M.D.V.; ASSIS, T.F. Seleção recorrente recíproca entre populações sintéticas multi-espécies (SRR-PSME) de *Eucalyptus*. **Pesquisa Florestal Brasileira**, v.57, p.57-60, 2008.

RESENDE, M.D.V.; BARBOSA, M.H.P. Melhoramento genético de plantas de propagação assexuada. Colombo: **Embrapa Florestas**, 2005. 130p.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L; PIRES, I.E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal**, n. 56, 2008.

RESENDE, M.D.V.; RESENDE, M.F.R.; SANSALONI, C.P. MISSIAGGIA AA, AGUIAR A.M., ABAD J.M., TAKAHASHI E.K., ROSADO A.M., FARIA D.A. et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, n.1, p.116-128, 2012.

REZENDE, B.A. seleção genômica ampla para volume e qualidade da madeira em *Eucalyptus*. 2015, 74p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras.

ROSS, R.J.; BRASHAW, B.K.; PELLERIN, R.F. Nondestructive evaluation of wood. **Forest Products Journal**, v. 48, n. 1, p.14-19, 1998.

ROSSO, S.; MUNIZ, G.I.B; MATOS, J.L.M.; HASELEIN, C.R.; HEIN, P.R.G, LOPES, M.C. Density of *Eucalyptus grandis* W. Hill Ex Maiden using Near Infrared Spectroscopy. *Cerne*, Lavras, v. 19, n. 4, p. 647-652, 2013.

SAMISTRARO, G.; DE MUNIZ, G; G. I. B.; PERALTA-ZAMORA, P.; CORDEIRO, G.A. Previsão das propriedades físicas do papel Kraft por espectroscopia no infravermelho próximo (NIR) e regressão por mínimos quadrados parciais (PLS). *Química Nova*, v.32, n. 6, p.1422-1425, 2009.

SANTOS, R.B.; GOMIDE, J.L.; SOUSA, L.C. Prediction of wood and Kraft Pulp quality by Near Infrared Spectroscopy. *R. Árvore*, Viçosa-MG, v.33, n.4, p.759-767, 2009.

SCHAEFFER, L.R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 123, n.4, p. 218-223, 2006.

SCHIMLECK, L.R., PAYNE, P.; WEARNE, R.H. Determination of important pulp properties of hybrid poplar by near infrared spectroscopy. **Wood and Fiber Science**, v.37, n. 3, p. 462-471, 2005.

SOLBERG, T.R.; SONESSON, A.; WOOLIAMs, J.; MEUWISSEN, T.H.E. Genomic selection using different marker types and density. **Journal of Animal Science**, v.86, n.10, 2447-2454, 2008.

SOUSA, L.C.; SANTOS, R. B.; GOMIDE, J.L. Prediction of wood and kraft pulp quality by Near Infrared Spectroscopy (NIR). **Revista Árvore**. v.33, p. 759-767, 2009.

THAMARUS, K.; GROOM, K.; BRADLEY, A.; RAYMOND, C.A.; SCHIMLECK, L.R.; WILLIAMS, E.R.; MORAN, G.F. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theor Appl Genet* v. 109, p. 856-864, 2004.

VIANA, A. P. ; RESENDE, M.D.V. DE . *Genética Quantitativa no Melhoramento de Fruteiras*. 1. ed. Rio de Janeiro: Interciência, v. 1. 284p, 2014.

VIANA, L.C.; TRUGILHO, P.F.; HEIN, P.R.G.; SILVA, J.M; LIMA, J.T. Calibration models and near infrared spectroscopy for predicting chemical properties and wood basic density in *Eucalyptus*. **Ciência Florestal**, v. 20, p. 2, 2010.

WONG, C.K.; BERNARDO, R. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, v.116, n.6, p. 815-824, 2008.

ZHONG, S.;DEKKERS, J.C.M.;FERNANDO, R.L.;JANNINK, J-L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, n.1, p. 355-364, 2009.

CAPÍTULO I - CAPACIDADE PREDITIVA DO MODELO DE SELEÇÃO GENÔMICA EM *Eucalyptus benthamii* PARA VOLUME E QUALIDADE DA MADEIRA¹

RESUMO

O objetivo desse trabalho foi avaliar a capacidade preditiva da Seleção Genômica para caracteres de crescimento e qualidade da madeira. Para o desenvolvimento desse trabalho foi realizada a fenotipagem da população de ‘treinamento’ aos 3 anos de idade para volume de madeira, em seguida a genotipagem utilizando marcadores moleculares do tipo SNPs nos 780 indivíduos selecionados por meio de modelos mistos. Essa genotipagem foi realizada utilizando um chip de alta densidade que resultou em 43.266 marcadores SNPs, após o controle de qualidade 27.462 marcas foram utilizadas para a estimativa dos efeitos dos marcadores. Também foi caracterizada a madeira aos 4 anos: densidade básica da madeira, lignina, extrativos e carboidratos. Foram realizados três tipos de validações cruzadas por meio da reamostragem de um grupo de indivíduos via procedimento Jackknife. A predição de valores genéticos genômicos foi obtida para todos os caracteres avaliados utilizando o modelo RR-BLUP, sendo que a capacidade preditiva variou de 0,24 a 0,74 para volume de madeira aos três anos de idade. Para densidade básica da madeira, teor de lignina, extrativos e carboidratos, as capacidades preditivas foram 0,44, 0,39, 0,33 e 0,35 respectivamente. Conclui-se que a seleção genômica ampla é aplicável para a espécie *Eucalyptus benthamii*. Com 2 a 5 mil marcadores consegue-se acessar a variabilidade para as características de crescimento e qualidade da madeira.

Palavras-chave: Genotipagem, marcadores SNPs, seleção de marcas

¹ As normas desse capítulo seguem às da Revista Tree Genetics and Genomes

**PREDICTIVE ABILITY OF THE GENOMIC SELECTION MODEL IN
Eucalyptus benthamii FOR WOOD VOLUME AND WOOD QUALITY**

ABSTRACT

The aim this work was to evaluate predictive ability of the genomic selection for growth and wood quality. To develop this work was done the phenotyping of the training population at three years old for wood volume; next the genotyped was done using SNPs molecular markers to 780 selected individuals by traditional BLUP. The genotyping was performed using a high density chip which resulted in 43,266 SNPs markers, after quality control marks 27,462 were used to estimate the effects of markers. The phenotyping was made for wood traits at four years old: wood basic density, lignin, extractives and carbohydrates. Three types of cross-validations were performed by resampling a group of individuals via the Jackknife procedure. The prediction of genomic breeding values was obtained for the wood volume trait using the RR-BLUP model, and the predictive ability ranged from 0.24 to 0.74 for wood volume at three years of age. For basic wood density, lignin content, extractives and carbohydrates, the predictive capacities were 0.44, 0.39, 0.33 and 0.35 respectively. It's been that the genomic selection is applicable for the species *Eucalyptus benthamii*. With 2 to 5 thousand markers it is possible to access the variability for the traits of growth and wood quality.

Keywords: Genotyping, SNPs, marks selection.

1. INTRODUÇÃO

O Brasil é referencia na silvicultura do *Eucalyptus* e esse mérito foi em função dos grandes avanços no melhoramento genético tradicional, bem como em melhores práticas de manejo silviculturais desenvolvidas. Da área plantada com *Eucalyptus* no Brasil 11,6% está concentrado na região Sul (IBÁ, 2016). O *Eucalyptus benthamii* é uma espécie de grande importância no sul do Brasil devido sua alta tolerância ao frio aliado à boa adaptação e crescimento volumétrico. Porém, essa espécie de origem Australiana possui pequena região de ocorrência nesse país com uma base genética mais estreita comparada a outras espécies de eucalipto introduzidas no Brasil. As populações introduzidas dessa espécie no Brasil são provenientes de poucas famílias. O melhoramento genético dessa espécie é bem recente e a principal estratégia usada atualmente é a Seleção Recorrente Intrapopulacional (SRI). A seleção de indivíduos superiores nos programas de melhoramento genético tem sido praticada pelo valor genotípico estimado a partir de dados fenotípicos de poucas características mensuradas em campo, sendo o tempo um dos maiores gargalos na seleção podendo levar em alguns casos até 15 anos para o avanço de uma geração.

Para otimizar e acelerar os ganhos genéticos nesses programas foi proposta e desenvolvida na década de 1990 a seleção assistida por marcadores (SAM) (LANDE & THOMPSON, 1990), foram realizados trabalhos para diversas culturas ao longo desses anos, inclusive espécies florestais (JUNGHANS et al., 2003; THAMARUS et al., 2004; MISSIAGIA et al., 2005), porém a implementação da SAM foi limitada e os ganhos em eficiência reduzidos. O grande problema é que um marcador para ser útil precisa explicar grande parte da variação genética de uma característica quantitativa, mas essas são controladas por vários genes de pequenos efeitos e, geralmente encontrados um

pequeno número de QTLs de grande efeito, os quais não explicam suficientemente toda a variação genética (DEKKERS, 2004; BERNARDO, 2008).

Como alternativa à SAM foi proposta em 2001 a Seleção Genômica (SG) que visa acelerar os ganhos genéticos por ciclos de seleção (MEUWISSEN et al., 2001). A SG é denominada ampla porque atua em todo o genoma sem a necessidade prévia de identificar marcadores moleculares com efeitos significativos e de mapear QTLs. Essa pode ser aplicada em todas as famílias em avaliação nos programas de melhoramento genético de espécies alógamas, apresenta alta acurácia seletiva para a seleção baseada exclusivamente em marcadores e não exige prévio conhecimento das posições (mapa) dos QTLs, não estando sujeita aos erros tipo II de aceitar uma hipótese verdadeira como falsa.

A utilização da seleção genômica é recente para melhoramento de espécies arbóreas e ainda está em fase de validação, mas essa técnica tem demonstrado e confirmado superioridade e exequibilidade prática em benefício do melhoramento animal (DEKKERS, 2012), o que é positivo, visto que o BLUP e VGG foram derivados do melhoramento animal, o que provavelmente trará um rápido reconhecimento da seleção genômica no melhoramento de árvores (GRATTAPAGLIA, 2014).

Com o desenvolvimento e total validação da SG será possível excluir algumas etapas dos programas de melhoramento, como teste de progênie, teste clonal e, dessa maneira, os futuros clones serão recomendados com um tempo mínimo para a avaliação em campo quanto à competição intra-específica, denominados plantios pilotos ou testes clonais ampliados. Essa inovação tecnológica será muito importante para o setor florestal, pois poderá acelerar os ganhos de seleção e aumentar ainda mais a competitividade brasileira nesse setor (GRATTAPAGLIA, 2014).

Os primeiros trabalhos a demonstrar a aplicabilidade da seleção genômica no

melhoramento florestal foram os de RESENDE et al. (2008) e GRATTAPAGLIA & RESENDE (2011), ambos usando simulação e RESENDE et al. (2012) em *Eucalyptus* (RESENDE, M.F.JR. et al., 2012a) em *Pinus*.

Em *Eucalyptus*, os primeiros estudos de prova de conceito da SG foram realizados por RESENDE et al., (2010) e RESENDE et al., (2012). Nesse trabalho foram utilizados 3000 marcadores dominantes do tipo DArT para duas populações distintas, com diferentes tamanhos efetivos e avaliou as características altura, diâmetro e penetração do pilodyn, sendo os resultados promissores, mostrando que a SG é factível em *Eucalyptus*. Um estudo mais amplo, por Lima (2014), possibilitou maior acesso aos efeitos de marcadores utilizando um chip de marcadores de alta densidade contendo 60.639 SNPs, desenvolvido para espécies de *Eucalyptus* (SILVA-JÚNIOR et al., 2015). Os modelos preditivos de seleção genômica, para volume e características da madeira, foram construídos baseados em duas metodologias RR-BLUP e BLASSO resultando em boas capacidades preditivas, porém sem diferença significativa entre os modelos, além de confirmar a capacidade preditiva da seleção genômica (acurácia de 0,53 para volume de madeira), foi possível verificar que o ranking da seleção assistida por marcadores coincidia em 77% com o de seleção fenotípica tradicional para o volume, o que confirma a aplicação prática de seleção genômica na população estudada (LIMA, 2014).

O objetivo desse trabalho foi avaliar capacidade preditiva da seleção genômica para volume e características físico-químicas da madeira em *Eucalyptus benthamii*.

2. MATERIAL E MÉTODOS

2.1 FENOTIPAGEM

2.1.1 Fenotipagem para volume de madeira

A população usada na estimação dos efeitos genéticos e validação dos marcadores foi composta por um teste de progênie de polinização aberta. Esse teste foi plantado em outubro de 2010, no município de Otacílio Costa-SC, em delineamento de blocos ao acaso, sendo 81 famílias com 20 repetições e uma planta por parcela, totalizando 1620 indivíduos. As sementes utilizadas nesse teste são provenientes de uma população base de 32 famílias procedentes da Austrália.

Todos os indivíduos foram fenotipados aos 3 anos de idade para DAP (Diâmetro a Altura do Peito) e altura, com isto foi possível estimar o volume de madeira aos 3 anos (VOL3). Para compor a população de treinamento, foram selecionados 780 indivíduos nesse teste de progênie para a característica volume de madeira, para isso, foi utilizado modelo misto REML/BLUP (RESENDE, 2007a):

$$y = Xr + Za + e$$

Em que y é o vetor de dados, r é o vetor dos efeitos de repetição (assumidos como fixos) somados à média geral, a é o vetor dos efeitos genéticos aditivos individuais (assumidos como aleatórios), e é o vetor de erros ou resíduos (aleatórios). X e Z representam as matrizes de incidência para os referidos efeitos.

2.1.2 Fenotipagem para caracteres de qualidade da madeira

Foi realizada a coleta não destrutiva nas 780 árvores selecionadas com um trado de incremento de 12 mm para a retirada de 2 baguetas na altura do DAP (Diâmetro a Altura do Peito), uma bagueta foi para determinação da densidade e outra para caracterização química.

Posteriormente as baguetas da caracterização química foram transformadas em serragem, utilizando-se moinho Wiley e encaminhadas para as análises no laboratório de Celulose e Papel na UFV. Essas amostras foram analisadas quimicamente visando estabelecer características tecnológicas para produção de polpa celulósica: com base no teor de lignina, extrativos e carboidratos (soma dos polímeros que compõe celulose: glicanas e hemiceluloses: xilanas, arabinanas, galactanas e mananas).

As serragens já processadas e moídas foram classificadas em peneiras de malhas 40 e 60 mesh, acondicionadas em sala com umidade relativa e temperaturas constante (UR 50% e 23°C) e, posteriormente, armazenadas em frascos hermeticamente fechados. As análises químicas foram realizadas de acordo com as normas apresentadas na Tabela 1.

Tabela 1. Análises químicas da madeira e normas utilizadas para determinação.

Características analisada	Metodologia
Extrativos - solubilidade em acetona	TAPPI T280 PM-99 (2000)
Lignina solúvel	GOLDSCHMID (1971)
Lignina insolúvel	GOMIDE & DEMUNER (1986)
Análise de carboidratos	WALLIS et al. (1996)

A densidade básica da madeira (DBM) foi determinada a partir do segundo conjunto de baguetas que foi retirada na altura do DAP, utilizou-se o método da balança hidrostática, segundo norma da ABNT- Associação Brasileira de Normas Técnicas, pela relação entre o peso absolutamente seco e o volume saturado da amostra, segundo a Equação 1:

$$DBM = M_s/V_u \quad (1)$$

onde:

DBM é a densidade básica da madeira (g cm⁻³),

M_s é a massa seca da bagueta (g) em estufa (105±1 °C) e

V_u é o volume saturado da bagueta (cm³).

Após fenotipagem das características da madeira utilizada nesse capítulo para a construção dos modelos de predição via seleção genômica, os dados também foram utilizados para a construção dos modelos de predição via espectroscopia de infravermelho próximo (capítulo II).

A partir desses modelos foram realizadas as predições para as características de qualidade da madeira dos indivíduos não selecionados para a população de estimação, assim, em torno de 367 indivíduos tiveram seus valores de densidade básica da madeira, lignina, extrativos e carboidratos preditos via espectroscopia NIR. De posse da caracterização completa da população em estudo (780 selecionados + 367 indivíduos não selecionados) para as características de qualidade da madeira, foram determinadas as estimativas de componentes de variância e parâmetros genéticos pelo método da máxima verossimilhança restrita e melhor predição linear não viciada, fazendo uso do software SELEGEN-REML/BLUP (RESENDE, 2007a), a partir do seguinte modelo estatístico, $y = Xr + Za + e$, já descrito anteriormente.

Para estimativa das acurácias foi utilizado as estimativa das herdabilidades moleculares conforme equação proposta por Resende (2014).

$$h_M^2 = \frac{\sum_i^{nM} 2 \cdot p_i \cdot q_i \cdot m_i^2}{\sigma_y^2}$$

Em que:

h^2_M = herdabilidade molecular

n_M = número de marcas

σ^2_y = variância fenotípica

p_i = frequência de um dos alelos do loco marcador i

q_i = frequência do outro alelo do loco marcador i

m_i = efeito genético de um dos alelos do loco marcador i

A capacidade preditiva foi estimada a partir da correlação dos valores genéticos genômicos e os valores genéticos dos indivíduos em estudo e a acurácia de acordo com a equação de Resende (2014).

$$rgg^* = (rgg) / \sqrt{(h^2_t)} \cdot \sqrt{(h^2_M)}$$

Em que:

rgg^* = acurácia preditiva

rgf = capacidade preditiva

h^2_M = herdabilidade molecular

h^2_t = herdabilidade individual média de famílias

2.2 GENOTIPAGEM

O DNA foi extraído de folhas das 780 plantas selecionadas, essa extração foi realizada pela empresa Heréditas utilizando o método CTAB (GRATTAPAGLIA; SEDEROFF, 1994) e quantificados em equipamento nanodrop para concentrações finais entre 20-40 ng.μL⁻¹. As amostras de DNA foram genotipadas usando um chip de 60.639 SNPs, EuchIP60K específico para *Eucalyptus spp.* desenvolvido por SILVA JUNIOR et al. (2015). A construção desse chip de marcadores SNP foi realizada com

base nos parâmetros fornecidos pelo fabricante (Illumina Assay design Tool), buscando manter a distribuição uniforme prevista dos SNPs ao longo do genoma.

A genotipagem resultou em 43.266 SNPs, eliminados marcadores com quantidade de valores perdidos call rate >90%, para retirar alelos pouco informativos, ou seja, de baixa variação na população, foi utilizado MAF (minor allele frequency) >0.01. Este controle de qualidade resultou em 27.462 SNP. Foi realizada a imputação com 2p para dados perdidos. A densidade de marcas foi alta ao longo do genoma sendo em média 1 SNP cada 22 kbp.

2.3 ANÁLISE GENÔMICA

As capacidades preditivas para volume e características de qualidade de madeira foram determinadas pelo método de estimação dos efeitos de marcadores proposto por MEUWISSEN (2001) utilizando o método RR-BLUP (Random Regression Best Linear Unbiased Predictor), para isso foi utilizado o seguinte modelo linear misto geral para estimar os efeitos de marcadores:

$$y = Xb + Zm + e$$

$$m \sim N(0, I\sigma^2_m)$$

em que: y é o vetor de observações fenotípicas, b é o vetor de efeitos fixos, m é o vetor dos efeitos aleatórios de marcadores e refere-se ao vetor de erros aleatórios.

X e Z são as matrizes de incidência para b e h .

A matriz de incidência Z contém os valores 0, 1 e 2 para o número de alelos (do SNP) ou marcadores do tipo h_i no indivíduo diplóide j .

As equações genômicas de modelo misto para a predição de m via o método RR-BLUP equivalem a:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \frac{\sigma_e^2}{(\sigma_g^2/n)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

em que σ_g^2 refere-se à variância genética total do caráter σ_e^2 é a variância residual e n é o número de QTLs que pode ser estimado via $n_Q = \sum_{j=1}^n 2p_j(1-p_j)$, em n_Q é o número de locos marcadores, em que p_j é a frequência de um marcador o loco j.

O valor genético genômico global do indivíduo j é dado por:

$$VGG = \hat{y}_j = \sum_i Z_{ij} \hat{m}_i$$

Em que o componente Z_{ij} refere-se ao elemento i da linha j da matriz Z, referente ao indivíduo, em que Z_i equivale a 0, 1 ou 2.

As equações de predição apresentadas anteriormente assumem, *a priori*, que todos os locos explicam iguais quantidades da variação genética. Assim, a variação genética explicada por cada loco é dada por (σ_g^2/n) , em que σ_g^2 é a variação genética total e n é o número de marcadores. A variação genotípica σ_g^2 pode ser estimada por REML sobre os dados fenotípicos da maneira tradicional ou pela própria variação entre marcadores ou dos segmentos cromossômicos de QTL.

No procedimento do RR-BLUP não há necessidade de uso da matriz de parentesco, baseada em pedigree. Essa é substituída por uma matriz de parentesco estimada pelos marcadores (RESENDE, 2007b).

Para estimar o efeito dos marcadores, a validação cruzada foi realizada por meio da reamostragem de um grupo de indivíduos via procedimento Jackknife, divididos em

um conjunto de dados amostrais (N) em (G) grupos de tamanho igual a (k), de forma que $N=G.k$, foram testadas três tipos de validações cruzadas para volume de madeira aos 3 anos de idade. Na divisão dos conjuntos dos dados amostrais, a mesma população foi utilizada em todas as validações.

No primeiro tipo de validação, a população em estudo foi dividida em 78 grupos de 10 indivíduos de maneira que a mesma análise foi repetida 78 vezes, porém em cada repetição um grupo foi removido da população e utilizado para a formação da população de validação e os outros 770 indivíduos foram utilizados na população de estimação dos efeitos de marcadores, foi realizada uma análise de efeitos dos marcadores, assim, os efeitos dos marcadores foram ranqueados em valor absoluto, formando-se grupos compostos por 1.000, 2.000, até 27.000 marcadores de maior efeito.

Para o segundo tipo de validação, a população em estudo foi dividida em 10 grupos de 78 indivíduos de maneira que a mesma análise foi repetida 10 vezes, enquanto os 78 indivíduos separados formavam a população de validação a cada ciclo, os outros 702 indivíduos reservados eram utilizados na população de estimação dos efeitos de marcadores, neste caso também, formou-se grupos de mil em mil com marcadores de maior efeito, de forma cumulativa até 27.000.

Na terceira validação realizada, foram formados 78 grupos de 10 indivíduos, sendo que a cada ciclo, diferente das validações anteriores, selecionou-se os marcadores de maior efeito realizando o ranqueamento do valor absoluto de mil em mil de forma cumulativa a cada vez que o ciclo acontecia. A cada ciclo a população de validação era composta por 78 indivíduos enquanto os outros 702 indivíduos foram utilizados na população de estimação dos efeitos de marcadores.

Foram realizadas as validações (Tabela 2) para as características de volume e

qualidade da madeira, aqui selecionou-se os marcadores de maior efeito realizando o ranqueamento do valor absoluto de mil em mil de forma cumulativa a cada vez que o ciclo acontecia.

Tabela 2. Descrição dos ciclos e grupos para a validação cruzada via procedimento Jackknife para características genotipadas e fenotipadas.

<i>Características</i>	<i>Nº marcadores</i>	<i>Nº de indivíduos</i>	<i>Nº de indivíduos/ciclo</i>	<i>Nº de ciclos</i>
DBM	27.462	720	80	9
Extrativos	27.462	738	82	9
Lignina	27.462	738	82	9
Carboidratos	27.462	740	74	10

A seleção genômica consiste em utilizar uma população para estimar o efeito dos marcadores e uma população de validação, como descrito acima, para analisar a estimação desses efeitos na recuperação do valor genômico de uma população independente. Após a validação e estimação de um conjunto de marcadores, o modelo poderá ser usado em uma população relacionada, onde será realizada a genotipagem dos indivíduos para posterior predição do seu fenótipo (RESENDE, 2007b).

3. RESULTADO E DISCUSSÃO

3.1 CARACTERIZAÇÃO DA POPULAÇÃO EM ESTUDO

As estimativas de herdabilidades para volume e qualidade da madeira das plantas foram apresentadas na Tabela 3. A média geral e as herdabilidades (h^2_a , h^2_{mp} , h^2_{ad}) para volume de madeira estão entre os valores encontrados na literatura para volume de madeira em *Eucalyptus spp.* (COSTA et al., 2016; ROSADO et al., 2009, HENRIQUES et al., 2017). A herdabilidade média de progênes foi alta (0,6) e a herdabilidade individual no sentido restrito foi moderada (0,2) evidenciando possibilidades de sucesso com a seleção.

As características da madeira apresentaram valores maiores de herdabilidades, sugerindo que estes caracteres sofrem menor influência ambiental. Para densidade básica da madeira aos 3 anos, HENRIQUES et al. (2017) encontrou 0,59 de herdabilidade individual e herdabilidade média de progênie de 0,82, bem próximo aos valores encontrados nesse trabalho, 0,68 e 0,80 (Tabela 3), respectivamente.

A acurácia na seleção progênes foi 77 % para o caráter volume de madeira, indicando alta precisão no acesso à variação genética verdadeira a partir da variação fenotípica observada nos caracteres. Para as características da madeira esses valores de acurácia foram bem superiores (84 a 90%) indicando maior confiança nos valores genéticos individuais.

Com a seleção dos 780 indivíduos para compor a população de treinamento para estudos de seleção genômica ampla o tamanho efetivo foi de 117.

Tabela 3. Parâmetros genéticos da população de *E. benthamii* para características de crescimento e qualidade da madeira.

Característica	Vol3	DBM	Extrativos	Lignina	Carboidratos
N	1348	1115	1115	1202	1117
h^2a	0,207	0,686	0,716	0,454	0,616
h^2mp	0,593	0,805	0,813	0,719	0,785
h^2ad	0,210	0,621	0,654	0,384	0,546
Acprog	0,772	0,897	0,902	0,848	0,886
Média	0,09	0,37	2,11	31,69	61,44

N = número de indivíduos avaliados para cada característica; h^2a = h^2 : herdabilidade individual no sentido restrito, ou seja, dos efeitos aditivos; h^2mp : herdabilidade da média de progênie, assumindo sobrevivência completa. Acprog: acurácia da seleção de progênie, assumindo sobrevivência completa. h^2ad : herdabilidade aditiva dentro de progênie.

3.2 PREDIÇÃO GENÔMICA

Predições genômicas foram realizadas para todas as características de crescimento e qualidade da madeira por meio do modelo RRBLUP. Para volume de madeira aos 3 anos de idade, a herdabilidade, a capacidade preditiva e acurácia foram calculadas a cada mil marcadores de forma cumulativa até o total geral que são 27.462 marcadores SNPs (Tabela 4). Com base nesses resultados, verificou-se que a maior valor de capacidade preditiva (0,74) foi alcançado com 5 mil marcadores, ou seja, um número menor de marcadores foi capaz de capturar variações genéticas para a característica volume de madeira aos 3 anos. Ao utilizar o estimador de acurácia (rgg*) proposto por RESENDE et al. (2014), obtiveram-se valores mais conservadores variando de 0,33 a 0,52, sendo que os valores dessas estimativas mantiveram dentro do espaço paramétrico (0 e 1). Esses valores de acurácia são mais conservadores, porém estes levam em consideração a herdabilidade genômica, essa, como proporção de herdabilidade fenotípica, leva em conta a eficiência dos marcadores em capturar os

QTL, ou seja, considera o grau de imperfeição do desequilíbrio de ligação. No entanto, para comparação com outros trabalhos da literatura, serão apresentados nesse estudo, valores de capacidade preditiva (r_{gf}), principalmente.

Com o estimador tradicional ($r_{gg} = 0,24$ a $0,74$) que somente considera a herdabilidade fenotípica, o valor de capacidade preditiva $0,74$, por exemplo, com herdabilidade fenotípica de $0,20$ e herdabilidade molecular de $0,10$, teria o valor de acurácia preditiva de $1,65$, ou seja, fora do espaço paramétrico.

Tabela 4. Acurácia de seleção genômica (rgg*) calculada na população de validação associada à herdabilidade individual para o caráter volume de 0.2, considerando herdabilidade genômica (h^2_M) e correlação ou capacidade preditiva (rgf), com os marcadores de maiores efeitos ordenados de 1.000 em 1000 de forma cumulativa.

Nº SNPs ⁽¹⁾	h^2_M	Rgf	rgg*
1000	0,040	0,74	0,33
2000	0,070	0,75	0,44
3000	0,080	0,75	0,47
4000	0,090	0,74	0,50
5000	0,100	0,74	0,52
6000	0,090	0,72	0,48
7000	0,080	0,70	0,44
8000	0,070	0,69	0,41
9000	0,060	0,67	0,37
10000	0,050	0,65	0,33
11000	0,040	0,63	0,28
12000	0,040	0,62	0,28
13000	0,030	0,61	0,24
14000	0,030	0,59	0,23
15000	0,020	0,58	0,18
16000	0,020	0,56	0,18
17000	0,020	0,55	0,17
18000	0,018	0,54	0,16
19000	0,016	0,53	0,15
20000	0,014	0,52	0,14
21000	0,012	0,50	0,12
22000	0,010	0,42	0,09
23000	0,008	0,38	0,08
24000	0,006	0,35	0,06
25000	0,004	0,30	0,04
26000	0,002	0,25	0,03
27000	0,000	0,24	0,00
27.462	0,120	0,44	0,34

Nos diferentes tipos de validação cruzada testados, as capacidades preditivas para volume de madeira aos 3 anos variaram de 0,12 a 0,75 (Figura 2). As Três metodologias de validação testadas foram similares, inclusive podemos observar que houve uma queda da capacidade preditiva com aumento do número de marcadores em nessas validações. Não existe explicação na literatura para esse comportamento, no

entanto, alguns trabalhos mostram que podem existir fontes de erro e superestimava da capacidade preditiva, durante os processos de validação.

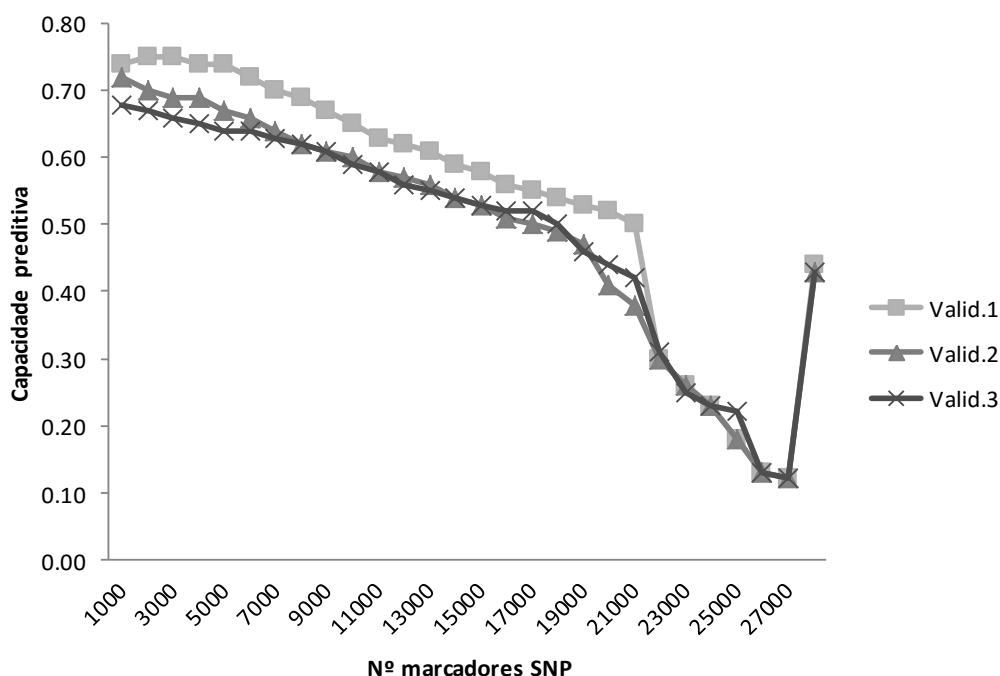


Figura 2 – Comparação da capacidade preditiva para volume de madeira aos 3 anos utilizando três métodos de validação. Validação 1- validação de 78 ciclos usando os efeitos médios estimados dos marcadores. Validação 2 - validação de 10 ciclos usando os efeitos médios estimados dos marcadores. Validação 3 - metodologia de validação de 78 ciclos usando marcadores de maiores efeitos.

Um erro que pode ocorrer nesses processos de validação é selecionar os SNPs mais significativamente associados na amostra inteira e usá-los para estimar os efeitos de SNP, ou seja, testar sua capacidade de predição nos conjuntos de descoberta e validação. Neste caso, a variância explicada pelos SNPs quando aplicada na amostra de validação é inflacionada. Ele cria *bias* e resultados enganosos porque a etapa de seleção inicial dos SNPs é baseada em uma correlação casual entre esses SNPs e toda a amostra, assim como entre os SNPs e qualquer subamostra. Uma equação de predição baseada nesses SNPs parecerá funcionar na amostra de validação, mas não em uma amostra genuinamente independente. Desta forma, a análise de validação cruzada após o

conjunto inicial de SNPs quando selecionada de toda a amostra não atenua esse viés. A armadilha da seleção de SNP a partir de amostras de descoberta e validação dependentes leva a erros (WRAY et al., 2013).

Desta forma, esse tipo de problema pode ter ocorrido nesse estudo provocando o decaimento da capacidade preditiva apresentado na Figura 2 e 3, visto que foi feita a seleção das marcas mais significativas usando toda a população para fazer a predição, ou seja, houve sobreposição entre as populações de treinamento e validação.

BERMINGHAM et al.(2015) trabalhando com a seleção genômica em humanos para as características altura, HDL e índice de massa corporal, também verificou que houve inflação das acurácias para todas as características quando a população de validação e estimação eram dependentes. Para resolver este problema foi feita a seleção de marcas de forma supervisionada usando somente os indivíduos da população de treinamento, deixando completamente de fora indivíduos a serem usados na validação.

Predições dos valores genéticos genômicos também foram obtidas para as características química e física da madeira utilizando o modelo RR-BLUP. Como pode ser observado na Figura 3, para todas as características da madeira também houveram decaimentos da capacidade preditiva à medida que se aumenta o número de marcadores SNPs.

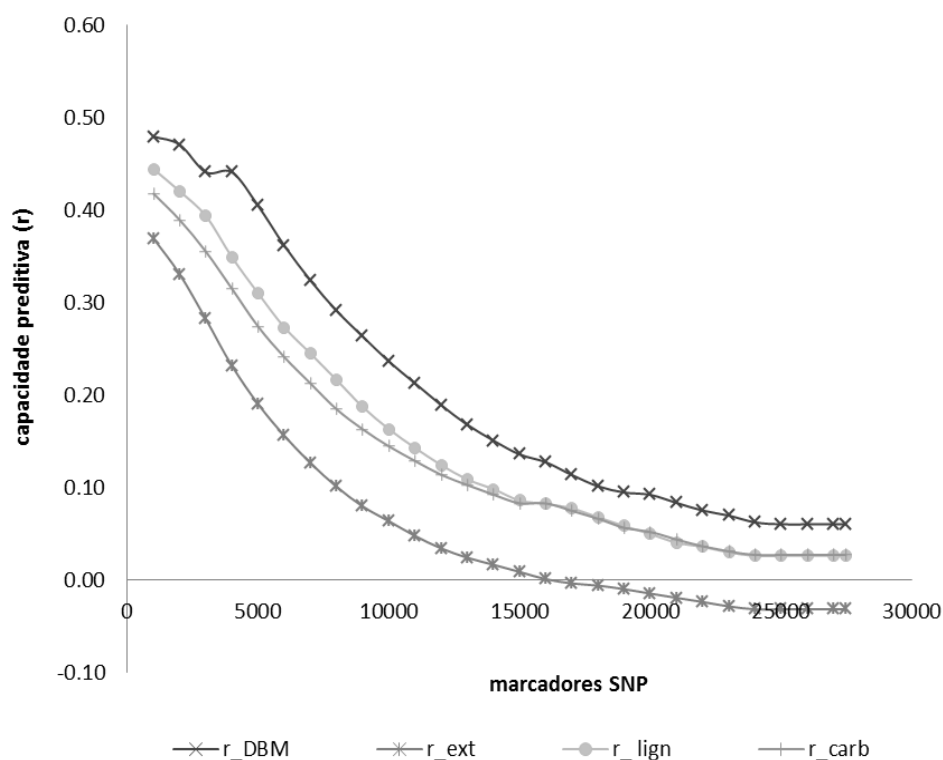


Figura 3. Capacidade preditiva (r) para características de densidade básica da madeira, extrativos, lignina e carboidratos aos 4 anos de idade para todos os marcadores SNP, variando de mil em mil.

Comportamento semelhante foi apresentado pelas características relacionadas à madeira, com um número menor de marcadores também foi possível acessar a variabilidade genética do *E. benthamii*, 2 a 3 mil foram suficientes, para volume de madeira 5 mil marcadores (Tabela 5). Porém, a utilização do chip com todos os marcadores moleculares poderá ser mais prudente nas próximas genotipagens, visto que esse foi o primeiro trabalho para essa população. Nas próximas genotipagens os modelos poderão ser retroalimentados e com o processamento da informação será possível entender melhor a população e esses problemas de estimativa.

Conforme apresentado na Tabela 5, as capacidades preditivas variaram de 0,30 a 0,74 e a herdabilidade molecular de 0,29 a 0,50, sendo que o *bias* (melhor próximo de 1) variou de 0,93 a 1,07, ou seja, ficou em torno de 1 para a maioria das características, nas quantidades de 2 a 5 mil marcadores. Isso mostra que os valores genéticos

genômicos preditos nessas quantidades de marcadores de maior efeito, não são viesados, por isso, são os pontos ótimos de predição. Como pode ser comparado, na Tabela 5, as acurácias com o estimador utilizado tradicionalmente podem variar de 0,38 a 1,65, sendo que as acurácias com o estimador que penaliza a estimativa pela herdabilidade molecular, variou de 0,42 a 0,52, ou seja, dentro do espaço paramétrico de 0 a 1.

Tabela 5. Ponto ótimo de capacidade preditiva dos marcadores SNP e as herdabilidades, acurácias e erros associados.

Característica	nº marcas	h^2	h^2_M	rgf	rgg	rgg*	By1
VOL3	5000	0,20	0,10	0,74	1,65	0,52	0,93
DBM	3000	0,68	0,50	0,44	0,53	0,51	1,01
Extrativos	2000	0,71	0,32	0,33	0,39	0,49	0,96
Lignina	3000	0,45	0,39	0,39	0,58	0,42	1,07
Carboidratos	3000	0,61	0,36	0,35	0,45	0,46	1,03

Herdabilidade genotípica (h^2); Herdabilidade genômica (h^2_M); Correlação ou capacidade preditiva (rgf); viés de predição (By1). Acurácia de seleção genômica considerando herdabilidade genômica (rgg*); Acurácia de seleção genômica considerando apenas herdabilidade genotípica (rgg).

Com um menor número de marcadores de maior efeito, para volume aos 3 anos em torno de 5 mil marcadores obteve-se a capacidade preditiva foi de 0,74 com herdabilidade de 0,20 e herdabilidade molecular de 0,10. As características da madeira apresentaram maiores herdabilidades, no entanto, isso não refletiu na melhoria da capacidade preditiva, como ocorreu em outros trabalhos onde a herdabilidade estava fortemente correlacionada com a capacidade preditiva (LIMA, 2014).

DBM e lignina apresentaram capacidade preditiva de 0,44 e 0,33, respectivamente. Para os carboidratos da madeira, que é a somatória da celulose e

hemicelulose da madeira, a capacidade preditiva foi de 0,35 com 3 mil marcadores, e herdabilidade de 0,61 (Tabela 5).

Em um dos primeiros trabalhos para *Eucalyptus spp.*, avaliou-se a eficácia da GS em duas populações não relacionadas, porém esses foram genotipados com 3000 marcadores 3000 DArTs. Uma população de 738 indivíduos híbridos de *E. grandis* x *E. urophylla*, obteve capacidades preditivas de 0,54 para circunferência e 0,51 para altura. Para a outra população de 920 indivíduos de *E. grandis*, *E. urophylla*, *E. globulus* e seus híbridos alcançaram valores de capacidades preditivas de 0,55 a 0,46 também para circunferência e altura (RESENDE et al., 2012).

Lima (2014) trabalhando com uma população elite de *Eucalyptus spp.*, também usando o chip EucHIP60k.br para 15 características de crescimento e qualidade da madeira, obteve capacidades preditivas bem superiores as encontradas nesse trabalho como foi apresentado na Tabela 5. A capacidade preditiva para volume de madeira foi de 0,42 e para lignina e densidade de madeira foi de 0,63. Encontrou para teor de celulose capacidade preditiva de 0,52 e herdabilidade de 0,72, e para hemicelulose, a acurácia foi de 0,57 com herdabilidade de 0,81.

TAN et al.(2016), trabalhando com duas espécies puras de *Eucalyptus spp.* e seus híbridos, também utilizando o chip de EucHIP60k.br para crescimento e características físico-químicas da madeira obtiveram valores de médios a baixos de capacidade preditiva quando usaram o modelo RR-BLUP. As capacidades preditivas para volume de madeira aos três anos variaram de 0,12 a 0,17, e densidade da madeira aos 4 anos de 0,42 a 0,47. Nesse estudo demonstraram que com o aumento da população de treinamento e aumento do número de marcadores e aumentava a capacidade preditiva. Para todas as características estudadas TAN et al. (2016), não houve melhoria na capacidade preditiva quando usaram mais de 5 mil SNPs.

Outros estudos tem demonstrado isso também (ZHANG et al., 2011; SU et al., 2012), porém nesse presente estudo, apesar de com poucos marcadores conseguirmos acessar a variabilidade do *E. benthamii*, 2 a 3 mil (Tabela 5), ocorreu queda das capacidades preditivas com números maiores de marcadores como já discutido anteriormente.

Em uma população de estimação híbrida de *Eucalyptus spp.* composta por 611 indivíduos, avaliou-se a acurácia preditiva para 20 características relacionadas a crescimento e qualidade da madeira via 36.812 SNPs e 15.742 DArTs. Os valores de acurácia preditiva encontrados foram baixos para os dois tipos de marcadores, durante a validação cruzada para a maioria das características. Para densidade da madeira a acurácia foi de 0,36 e a herdabilidade de 0,30, já para lignina a herdabilidade foi de 0,28 e acurácia na validação foi de 0,22 (REZENDE, 2015).

Para comparar a seleção genômica com a seleção genotípica tradicionalmente utilizada, foi feita uma análise de coincidência entre o ordenamento dos genótipos via BLUP e RRBLUP. Desta forma, pode-se constatar que as porcentagens de coincidência são baixas, variando de 5 a 26% para os 30 melhores indivíduos selecionados. Porém, quando se avalia a superioridade da SG em relação a tradicional em função da redução de tempo do melhoramento genético (Tabela 6), podemos verificar superioridades até de 447 % quando se realiza a seleção de forma muito precoce, ainda em fase de muda.

Tabela 6. Superioridade da seleção genômica ampla (AG - Acurácia Genômica RRBLUP) comparada com a acurácia da seleção fenotípica (AF - Acurácia Fenotípica BLUP) em função da redução de tempo do melhoramento tradicional convencional em anos (TF-Tempo Fenotípica) para tempos reduzidos (TG-Tempo Genômica) para características de qualidade da madeira.

Característica	AF	AG	TF	TG	EF	S%
VOL3	0,77	0,52	4,0	2,0	1,35	35,06
	0,77	0,52	4,0	1,0	2,70	170,13
	0,77	0,52	4,0	0,5	5,40	440,26
DBM	0,90	0,53	4,0	2,0	1,18	18,11
	0,90	0,53	4,0	1,0	2,36	136,23
	0,90	0,53	4,0	0,5	4,72	372,45
Extrativos	0,90	0,39	4,0	2,0	0,86	-13,52
	0,90	0,39	4,0	1,0	1,73	72,96
	0,90	0,39	4,0	0,5	3,46	245,92
Lignina	0,85	0,58	4,0	2,0	1,37	36,78
	0,85	0,58	4,0	1,0	2,74	173,55
	0,85	0,58	4,0	0,5	5,47	447,10
Carboidratos	0,89	0,45	4,0	2,0	1,02	1,61
	0,89	0,45	4,0	1,0	2,03	103,21
	0,89	0,45	4,0	0,5	4,06	306,42

¹ EF (Eficiência em unidade de tempo e acurácia genômica e fenotípica) e S% (Superioridade); $EF = AG * TF / AF * TG$; $^2S\% = (EF - 1) * 100$; VOL3= volume de madeira aos 3 anos de idade; DBM (Densidade Básica da Madeira).

3.2 EFEITOS DOS MARCADORES SNP

Para ilustrar como os efeitos dos marcadores estão distribuídos nos cromossomos, foram construídos gráficos (Manhattan Plots) dos efeitos de SNPs para as características lignina e densidade da madeira. Para densidade da madeira (Figura 4) e para lignina (Figura 5) o cromossomo 8 foi o que apresentou maior efeito de marcas, tanto para o ponto ótimo com 3 mil quanto para o total de 27.462 marcadores, isso era esperado, visto que o oitavo é o maior cromossomo. Para lignina os cromossomos 2 e 3 foram os que apresentaram maior número de marcas com efeito pronunciado.

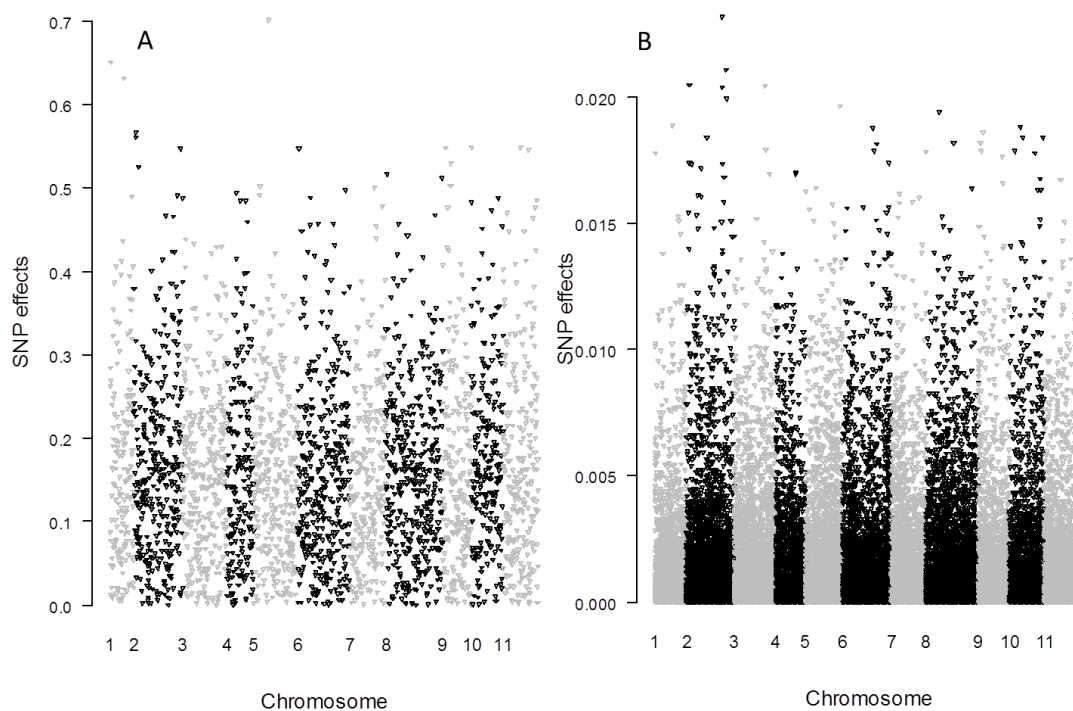


Figura 4 - Demonstração dos efeitos SNPs com três mil (A) e 27.462 marcadores (B) ao longo dos onze cromossomos do genoma de *E. benthamii* para densidade da madeira.

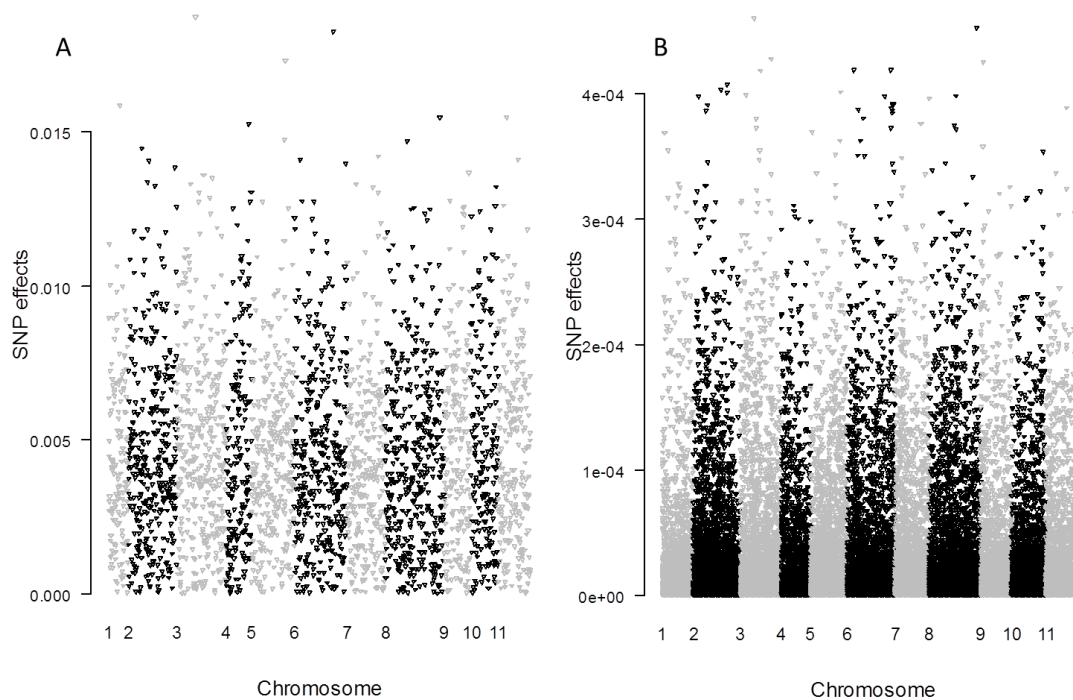


Figura 5 - Demonstração dos efeitos SNPs com três mil (A) e 27.462 marcadores (B) ao longo dos onze cromossomos do genoma de *E. benthamii* para teor de lignina total da madeira.

4. CONCLUSÕES

A seleção genômica é aplicável para volume e características físico-químicas da madeira em *Eucalyptus benthamii*.

Com 2 a 5 mil marcadores é possível acessar a variabilidade do *E. benthamii* para essas características. Porém, a utilização do chip com todos os marcadores moleculares poderá ser mais prudente nas próximas genotipagens, visto que esse foi o primeiro trabalho para essa população.

5. REFERÊNCIAS

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, Sept. 2008.

BERMINGHAM, M.L.; PONG-WONG, R., SPILIOPOULOU, A.; HAYWARD, C.; RUDAN, I.; CAMPBELL H.; WRIGHT, A.F.; WILSON, J.F.; AGAKOV, F.; NAVARRO, P.; HALEY, C.S. Application of high-dimensional feature selection: evaluation for genomic prediction in man. **Scientific Reports**. may, 2015. DOI: 10.1038/srep10312

COSTA, R.M.L.; ESTOPA, R.A.; BIERNASKI, F.A.; MORI, E.S. Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cambage por diferentes métodos de seleção. **Scientia Forestalis**, v. 45, n. 113, p. 105-113, 2017.

HENRIQUES, E.P; MORAES, C.B; SEBBENN, A.M.; TOMAZELLO FILHO, M.; MORAES, M.L.T; MORI, E.S. Estimativa de parâmetros genéticos para caracteres silviculturais e densidade do lenho em teste de progênies de *Eucalyptus urophylla*. **Scientia Forestalis**, v. 44, n. 109, p. 105-113, 2016.

DEKKERS, J. C. M. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, n.13, p.313-328, Jan. 2004.

DEKKERS, J. C. M. Application of Genomics Tools to Animal Breeding. **Current Genomics**, v.13, p. 207-212, 2012.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. (Ed.). **Genomics of plant genetic resources**. Dordrecht: Springerchap. 26, p. 651-682, 2014.

GRATTAPAGLIA D; RESENDE M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and Rapd markers. **Genetics**, Bethesda, v. 137, n. 4, p. 1121-1137, 1994.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 323-330, 2007.

GOLDSCHMID, O. Ultraviolet spectra. In: Sarkanen, K.V. e Ludwig, C.H. Ed. **Lignins**. New York: Wiley-Interscience, 1971. p.241-266.

GOMIDE, J.L. and DEMUNER, B.J. Determination of lignin content in woody material: modification Klason Method. **O Papel**, v. 47, p. 8, 1986.

IBÁ. **Indústria Brasileira de Árvores**, ano base 2015, Brasília, 2016.

JUNGHANS, D.T.; ALFENAS, A.C., BROMMONSCHENKEL, S.H., ODA, S., MELLO, E.J., GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of major gene with RAPD markers. **Theoretical and Applied Genetics**, v. 108, n. 1, p. 175-180, 2003.

LANDE, R.; THOMPSON, R. Efficiency of marker assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n. 3, p.743-756, 1990.

LIMA, B. M. Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data. 2014, 92p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura/USP, Piracicaba.

MISSIAGIA, A. A.; PIACEZZI, A. L.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics and Genomes**, Heidelberg, v. 1, n. 1, p. 79-84, Oct. 2005.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p.1819-1829, 2001.

RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007a. 561p.

RESENDE, M.D.V. Selegen-REML/BLUP : Sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Colombo: **Embrapa Florestas**, 2007b. 359p.

RESENDE, M.D.V.; ASSIS, T.F. Seleção recorrente recíproca entre populações sintéticas multi-espécies (SRR-PSME) de *Eucalyptus*. **Pesquisa Florestal Brasileira**, v.57, p.57-60, 2008.

RESENDE, M.D.V.; RESENDE, M.F.R.; SANSALONI, C.P. MISSIAGGIA AA, AGUIAR A.M., ABAD J.M., TAKAHASHI E.K., ROSADO A.M., FARIA D.A. et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, n.1, p.116-128, 2012.

RESENDE, M.F.JR; MUNOZ, P.; RESENDE, M.D.V.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M; JOKELA, E. J.; MARTIN, T. A.; PETER, G. F.; KIRST, M. Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.). **Genetics**, v.190, n.4, p. 1503-1510, 2012a.

RESENDE, M.D.V. DE; SILVA, F. F. E. ; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência**. 1. ed. Visconde do Rio Branco: Suprema, 2014, v. 1. 881p.

REZENDE, B.A. seleção genômica ampla para volume e qualidade da madeira em *Eucalyptus*. 2015, 74p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras.

ROSADO, A.M.; ROSADO, T.B.; RESENDE JUNIOR, M.F.R.; BHERING, L.L.; CRUZ, C.D. Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*. **Pesquisa Agropecuária Brasileira**, v.44, n.12, p.1653-1659, 2009.

SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**, London, v. 208, n. 3, p. 830-845, 2015.

SU, G; BRONDUM, RF; MA, P; GULDBRANDTSEN, B.; AAMAND, GR; LUND, MS. Comparison of genomic predictions using medium-density (similar to 54,000) and high-density (similar to 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. **J Dairy Sci**. v. 95, n.8, p.4657-4665, 2012.

Technical Association of the Pulp and Paper Industry-TAPPI. T280 pm-99 standard - Acetone extractives of wood and pulp. **TAPPI Press**, 2000.

THAMARUS, K.; GROOM, K.; BRADLEY, A.; RAYMOND, C.A.; SCHIMLECK, L.R.; WILLIAMS, E.R.; MORAN, G.F. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. *Theor Appl Genet* v. 109, p. 856-864, 2004.

ZHANG, Z., DING X, LIU J., ZHANG, Q., DE KONING, D.J. Accuracy of genomic prediction using 778 low-density marker panels. **J Dairy Sci**. v. 94, n.7, p. 3642-3650, 2011.

WRAY,N.R.; YANG, J.; HAYES, B.J.; PRICE, A.L.; GODDARD, M. E.; VISSCHER, P.M. Pitfalls of predicting complex traits from SNPs. *Nature Rev Genet.* v. 14, p.507–515. 2013.

WALLIS, A.F.A.; WEARNE, R.H. and WRIGHT, P.J. Analytical characteristics of plantation eucalypt woods relating to Kraft pulp yields. **Appita Journal** v. 49, p. 6, 1996.

WONG, C.K.; BERNARDO, R. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, v.116, n.6, p. 815-824, 2008.

CAPÍTULO II - NIR SPECTROSCOPIC MODELS FOR WOOD TRAITS IN BREEDING PROGRAMS OF *Eucalyptus benthamii*²

ABSTRACT

Wood characterization must be done in large populations of *Eucalyptus* breeding programs in order to efficiently select elite trees. In this study, *Eucalyptus benthamii* wood was non-destructively characterized and the performance of near infrared (NIR) spectroscopy in estimating wood basic density, lignin, extractive, glucose, xylan contents and total carbohydrates was evaluated. NIR models for wood traits were built from data 481 trees of an *E. benthamii* progeny test (4-year-old) managed for pulp cultivated in Santa Catarina state, Southern Brazil. Increment cores were sampled for chemical and physical characterization in laboratory, as well as for NIR spectroscopy analyses. Three 350 samples were selected from PCA (Principal Component Analysis) for model calibrations whereas 131 were reserved for independent test validation. The *E. benthamii* wood presented the standards required for Kraft pulp processing. The predictive NIR models showed satisfactory ability for estimating the chemical properties of wood. The prediction models for total lignin, extractive, xylan contents and total carbohydrates showed coefficients of determination of 0.53, 0.65, 0.36 and 0.53, with RPD (Ratio of Performance to Deviation) values for these traits ranging from 1.3 to 2.3. The predictive model for basic density and glucose presented low coefficient of determination (0.13 and 0.10). However, isn't possible to use these models for ranking in genetic selection because there isn't correlation. Thus, NIR spectroscopy can potentially be applied in breeding programs, as it enables an early, non-destructive selection of trees with adequate physical and chemical properties for pulp production.

Keywords: physicochemical traits, wood characterization non-destructively, wood quality.

² As normas desse capítulo seguem às da Revista Cerne.

1. INTRODUCTION

Brazil has become a reference in *Eucalyptus* forestry due to its great advances in traditional breeding and improvements in forestry management practices. The total *Eucalyptus* planted area in Brazil was 5.56 million hectares in 2016: 11.6% of this surface is concentrated in the South (IBÁ, 2016). *Eucalyptus benthamii* is a species of commercial importance in Southern Brazil because it is highly resistant to cold, and it also presents good adaptation and volumetric growth. The breeding of this species is very recent, with current strategies focused on selecting intraspecific and multi-species hybrids, i.e., hybrid individuals that combine several important features to increase yield, resistance to cold, and, in particular, the quality of wood suitable for pulp production. *Eucalyptus benthamii* is a species with a very narrow genetic basis; the little knowledge available stems from the fact that only a few individuals have been sampled (FONSECA et al., 2010; ALVES et al., 2011).

Evaluating wood quality in breeding populations can be challenging because it is not always possible to fell an elite tree, due to its genetic value in the breeding programme. Thus, non-destructive evaluation methods are required to enable the collection of a small wood sample without compromising the survival of the individual. Another important factor is wood assessment age, given that wood properties should be determined, ideally, after a tree has reached its harvest age or near this (PASQUINI et al., 2007), which hinders selection time and generation advancement. Hence, the minimum optimal assessment age that allows a good correlation with the quality of 7-year-old wood must be determined to enable earlier selection (SCHIMLECK et al., 2005).

Considering the need for early and non-destructive analysis, near infrared (NIR) spectroscopy – which has been used for several purposes in laboratory research and

industrial processes – has been pointed out as a quick and inexpensive method, which has greatly contributed to the chemical (RAYMOND; SCHIMLECK, 2002; SOUSA, 2008; VIANA et al., 2010; MILAGRES, 2013, RAMADEVI et al., 2016), physical analysis of major wood components (SCHIMLECK et al., 2005; HEIN et al., 2009, ROSOS et al., 2013), pulp yield (SANTOS et al., 2009) and even paper properties (SAMISTRARO et al., 2009). Another application of NIR in trees was to indicate the technology can be used as an effective tool to distinguish between pure pine species and suggest that it will also distinguish hybrids from their parents (ESPINOZA et al., 2012). OLIVEIRA et al. (2015), analyzing the spectra of timber of four species of trees was able to separate this species by NIR spectroscopy and Artificial Neural Networks. Of native wood charcoal, near infrared spectrometry has been used to distinguish species (DAVRIEUX et al. 2010, NISGOSKI et al. 2015).

In regard to tree selection in breeding programs based on NIR spectroscopy information, many studies have presented promising findings. GREAVES et al. (1996) evaluated the genetic control of NIR spectra recorded on wood powder of 588 *Eucalyptus nitens*, while HEIN and CHAIX (2014) have recorded NIR spectra on solid wood of *Eucalyptus urophylla* x *E grandis* hybrids clones. In both studies they have calculated the ratio between the genetic variance and phenotypic variance of the NIR absorbance values for each wavenumber producing broad-sense NIR spectral heritability estimates. They reported that some ranges of the spectra taken on the wood of 6-year *Eucalyptus* clones presented heritability estimates greater than 0.50, indicating that the breeder could select potential trees using the most heritable NIR signature bands (HEIN; CHAIX, 2014).

Thus, the aim of this study was to develop NIR-based models calibrations for prediction of wood density, glucose, xylan, lignin, extractive contents and total

carbohydrates in *Eucalyptus benthamii* wood in order to efficiently select potential trees.

2. MATERIALS AND METHODS

2.1 GENETIC MATERIAL

An open pollinated progeny test with 1620 individuals from 81 *E. benthamii* mother plants, at 4 years of age, in the municipality of Otacílio Costa (geographic coordinates (27° 28' 59" S, 50° 07' 19" W, 884 m), in Santa Catarina state, Brazil, was used for the wood characterization and NIR models development.

2.2 SAMPLE COLLECTION AND PREPARATION

Sample for NIR spectra: Non-destructive collection of 481 samples (Figure 6) was performed using a 12 mm increment borer to extract a bark-to-bark increment core at breast height. The cores were kept in a control room with 50% moisture and 23°C ($\pm 2^\circ\text{C}$) of temperature. They were ground using a Wiley mill, followed by a cyclone mill, to uniform particle size. Samples were stored for at least 48 hours at room temperature under controlled humidity prior to acquisition of NIR spectra. For that purpose, sawdust was manually placed in the spectrophotometer sample chamber. The spectrophotometer was calibrated with the equipment's internal reference standards.

Sample for chemical analysis: By PCA (Principal Component Analysis) were selected 481 samples, separated in two groups, one for calibration (350 samples) and other for external validation (131 samples). These select samples were sent for chemical analysis at the Pulp and Paper Laboratory in Viçosa Federal University.

Sample for wood basic density: The same time 481 trees were sampled again. Non-destructive collection of the 481 samples was performed using a 12 mm increment

borer to extract a bark-to-bark increment core at breast height. Samples were sent to the laboratory, for wood basic density determination in Klabin Laboratory.



Figure 6 - Samples being taken from the standing tree (A). Increment core (B) and saturated samples (C) for determination of basic density.

2.3 SPECTRAL READINGS

The NIR spectra were acquired using a Bruker TANGO FT-NIR (Bruker Optics Ettlingen Germany) in diffuse reflectance mode; 32 scans were recorded from 10,000 to 4,000 cm^{-1} with a resolution of 16 cm^{-1} . Two readings were performed per sample and the means of 64 scans were calculated.

2.4 CHEMICAL ANALYSIS OF WOOD

Wood samples were chemically analyzed to determine the lignin, extractive, glucose, xylan, arabinans, galactans and mannan contents. And the total carbohydrates were considered as the sum of glucose, xylan, arabinans, galactans and mannan

contents. Before analyses, the wood samples were ground and the sawdust samples were graded using 40 and 60 mesh sieves and stored in a room with constant temperature and humidity (23°C and 50%, respectively), followed by storage in hermetically sealed vials. Chemical analyses were performed according to the standard methods (GOMIDE; DEMUNER (1986); WALLIS et al., 1996).

2.5 DETERMINATION OF WOOD BASIC DENSITY

wood basic density was determined on increment cores using the hydrostatic balance method, according to an adaptation of the NBR 7190 (ABNT, 1997) for increment core. This parameter was calculated through the relationship between the absolute dry weight and the saturated volume of the samples.

2.6 MODEL CALIBRATION

Calibration and validation models were performed using the Unscrambler software (CAMO AS, Norway, version 10.2), using Principal Component Analysis (PCA) for selecting samples and Partial Least Squares (PLS) regression for establishing the relationships between NIR spectra and wood traits. Principal Component Analyses (PCA) were performed for selecting samples for calibration and validation sets. The samples were separated into two groups using the Kennard and Stone algorithm (KENNARD; STONE, 1969): 350 samples were used for calibration while 131 samples were used for external validations. Sample reflectance values acquired with the NIR spectrophotometer at various wavenumbers were associated with the concentrations of the chemical constituents of the samples.

The following selection criteria were adopted for the development of predictive calibration models: coefficient of determination of models (R^2); root mean square error

of cross validation (RMSECV) and of prediction (RMSEP); and the lowest possible adequate number of PLS factors for each model. The RMSE (root mean square error) was calculated according to the equation below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{n}}$$

Where: y_i is known value of the sample constituent interest i ; \hat{y}_i is the estimated value of the sample i by calibration; n is the number of samples used for the calibration development.

Ratio of Performance to Deviation ($RPD = SD / SECV$ or SEP) was used to evaluate the predictive models where, (i) standard error of cross-validation (SECV), (ii) standard error of prediction (SEP), (iii) standard deviation (SD). This ratio evaluates calibration accuracy to discern samples (WILLIAMS; SOBERING, 1993). High values RPD means that the error of trait estimation is smaller than the standard deviation of the samples. On another hand, low values indicate the inefficiency of the model. According to SCHIMLECK et al. (2003) RPD ratios around 1.5 are useful for screening and can be used in breeding programs, but values above 2 for RPD are ideal.

The original (untreated) NIR spectra and first-derivative NIR spectra (15-point filter and a second order polynomial) using the Savitzky-Golay method (SAVITZKY; GOLAY, 1997) were used for PLS regressions. Samples that visually differed from the remaining samples, with high leverage and high residual variance, were detected in graphical analysis as outliers and excluded from the calculations.

2.7 USE OF NIR MODELS IN THE BREEDING PROGRAM

To characterize all study population (1620 individuals), the NIR models were used for prediction wood basic density, extractives, lignin and carbohydrates.

After to characterize of physicochemical traits for this population, the SELEGEN-REML/BLUP (RESENDE, 2007) was used to determine the variance components and genetic parameters. The mixed linear model (model 19) was used: $y = Xr + Za + e$, where y is the data vector, r is the vector of the repetition effects (assumed as fixed) added to the general mean, a is the vector of the individual additive genetic effects (assumed to be random), and e is the vector of errors or residues (Random). X and Z represent the incidence matrices for the said effects.

3. RESULTS AND DISCUSSION

The progeny test in this study was conducted with four-year-old trees to evaluate wood and silvicultural traits. Individual volumes averaged 0.1 m³, which is consistent with the average yield for this age in areas with the occurrence of frost. In the same region, the yield from operational-scale clonal plantations is 42 m³/ha/year at 7 years of age. This yield is above the national average, currently at 39 m³/ha/year (IBÁ, 2015).

3.1 NIR SPECTRA OF WOOD

The scores from Principal Component Analysis display the variability of the NIR spectra among the *Eucalyptus* trees. The first principal component (PC1) accounted for 72% of the spectra variability while the PC2 accounted for 24% of the spectral variation, for a combined total of 96% of the spectral data variation of the samples. The first derivative NIR spectra ranging from 10,000 to 4,000 cm⁻¹ were recorded on increment cores of 350 trees.

3.2 NIR MODELS FOR WOOD TRAITS

3.2.1 Wood Basic Density

The model for wood basic density (BWD) was calibrated with the PLS method using 5 factors (Figure 7) and reached values of 0.37 and 0.026 g/cm³ for the coefficient of determination (R²) and RMSEC, respectively. Twenty groups were used for cross-validation, resulting in a coefficient of determination of 0.33 and RMSECV of 0.027 g/cm³ (Table 1). For external validation, 131 samples, that were not included in the model calibration, were used for validation, resulting in a coefficient of determination of 0.13, a RMSEP of 0.023 g/cm³ and, also presented almost acceptable RPD (1.4) value to

screening, but due the low error. FANTUZZI et al. (2011) achieved better results working with *Eucalyptus* spp., with the wood basic density yielding a coefficient of determination of 0.87 (error: 0.01) and a prediction coefficient of determination of 0.68 (error: 0.017 g/cm³). To examine the expected consistency of individual ranking we estimated Spearman's correlations between reference of BWD value and prediction value estimated by the NIR model. Such correlations can show how differently individuals are ranked in a selection stage during a breeding program. The Spearman's correlation was 0.34, considerably low.

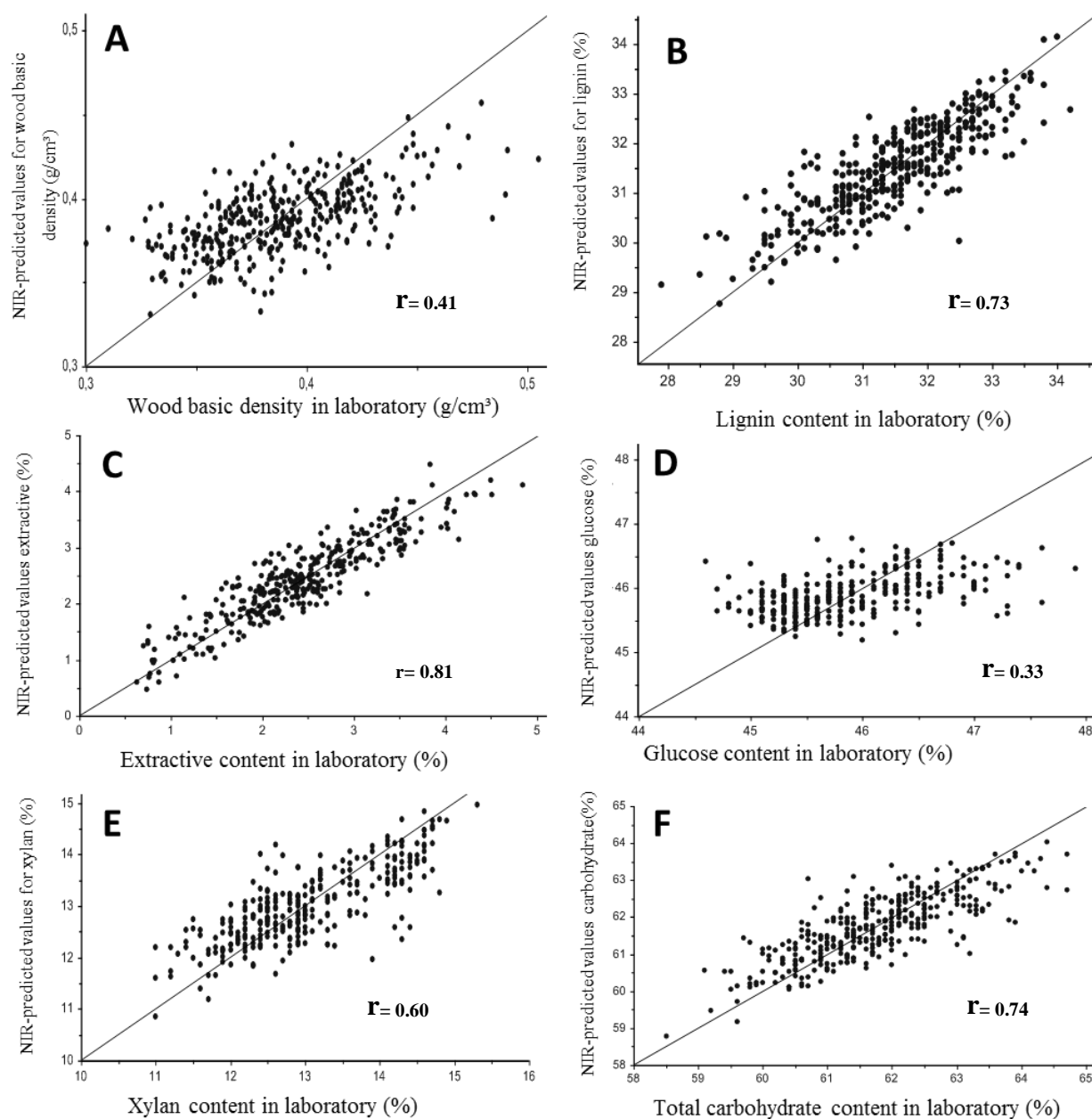


Figure 7 - Correlation of reference values versus predicted NIR for wood basic density(A), lignin (B), extractive (C), glucose (D), xylan (E) and total carbohydrate (F) for independent validation test set.

Table 1 - Calibration of NIR prediction models for basic density (g/cm³), total lignin (%), extractive (%), glucose (%), xylan content (%) and total of carbohydrates (%);

Wood traits	Model	# of samples	R ²	RMSE	# of factors	RPD
Basic density	Calibration	350	0.37	0.026	5	
	Cross-validation	20 groups	0.33	0.027	5	1.4
	External validation	131 (16 outliers)	0.13	0.023	5	
Lignin	Calibration	350 (344)	0.71	0.59	9	
	Cross-validation	20 groups	0.71	0.59	9	1.5
	External validation	131 (0 outliers)	0.53	0.77	9	
Extractives	Calibration	350 (344)	0.82	0.33	7	
	Cross-validation	20 groups	0.80	0.35	7	2.3
	External validation	131 (0 outliers)	0.65	0.34	7	
xylan	Calibration	300 (295)	0.61	0.57	5	
	Cross-validation	20 groups	0.57	0.60	5	1.3
	External validation	177 (0 outliers)	0.36	0.68	5	
glucose	Calibration	300 (295)	0.24	0.55	3	
	Cross-validation	20 groups	0.22	0.56	3	1.3
	External validation	177 (0 outliers)	0.10	0.59	3	
Carbohydrates	Calibration	350 (344)	0.65	0.66	6	
	Cross-validation	20 groups	0.69	0.61	6	1.7
	External validation	131 (0 outliers)	0.53	0.76	6	

R²: Coefficient of determination of models; RMSE: Root Mean Square Error; RPD: Ratio of Performance to standard Deviation.

To our knowledge, only one study has used NIR spectra to predict wood basic density in *E. benthamii* (LAZZAROTTO; MAGALHÃES, 2014): forty *E. benthamii* and 44 *E. pellita* trees were used to calibrate the model for wood basic density with coefficient of determination above 0.8 for the calibration, standard error of prediction of 0.027 g/cm³ and coefficient of determination in prediction set of 0.62. PASQUINI et al,

(2007), reported similar results in 926 *Eucalyptus* samples for model calibration and 126 samples for external validation. Their prediction error of wood basic density was 0.026 g/cm³ with R² was 0.86. However, SOUSA et al. (2011), working with *Eucalyptus spp.*, found a coefficient of determination of 0.89, but error of 0.034 g/cm³.

Thus, when comparing the wood basic density prediction errors found in the literature with the ones obtained in this study (0.023 g/cm³), we conclude that the error is low, but isn't possible to use this model to determine wood basic density for ranking in genetic selection because there isn't correlation. The best way is to feed this model with the next samples that are analyzed until the model reaches good determination coefficients to predict basic wood density.

3.2.2 Total Lignin

Calibration of total lignin via PLS regression resulted in a model with nine factors (Figure 7), and values of 0.71 and 0.59 for the coefficient of determination and RMSEC, respectively. Twenty groups were used for cross-validation, resulting in coefficient of determination of 0.71 and RMSECV of 0.59 (Table 1). External validation was performed using 131 samples not used for model calibration, resulting in a coefficient of determination of 0.53, RMSEP of 0.77 and presented acceptable RPD (1.5) value to screening.

MILAGRES et al. (2013) found a coefficient of determination of 0.55 to 0.71 and an error of 0.9 to 1.1 in cross validation for the total lignin content model using 136 *Eucalyptus* spp hybrid trees at the ages 3-7 years. In *E. urophylla*, HEIN et al. (2010) found a coefficient of determination of 0.89 and an error of 0.42 in the fit of the total lignin content model using 60 samples, along with a coefficient of determination of 0.84 and error of 0.5 in the cross-validation. The results of this study are similar to others in

the literature for the *Eucalyptus* genus, demonstrating the potential use of the lignin prediction model developed herein.

3.2.3 Extractives

The model for extractive content was calibrated using the PLS regression method with seven factors (Figure 7), yielding a coefficient of determination of 0.82 and an RMSEC of 0.33. Cross-validation was performed using 20 groups, resulting in a coefficient of determination of 0.80 and an RMSECV of 0.35 (Table 1). External validation was performed using 131 samples that were not included in model calibration, resulting in a coefficient of determination of 0.65, RMSEP of 0.34 and presented high RPD value (2.3) to estimate and screening,

The NIR-based model for extractives presented above generated the best prediction results, consistent with the majority of studies found in the literature. FANTUZZI et al. (2011) reported a coefficient of determination of 0.90 and an error of 0.4, as well as a prediction coefficient of determination of 0.84 with an error of 0.5. Meanwhile, SOUSA (2008) presented a low coefficient of determination (0.64) and an error of 0.65 in a calibration model for extractive contents in *Eucalyptus spp.* All studies involving NIR calibrations support the idea that the prediction model developed herein is adequate for predicting extractive content in wood, given that the prediction error (0.35) is lower than the values reported in other studies, and there was good correlation between the predicted value and the reference data (0.81).

3.2.4 Carbohydrates

The glucose model was calibrated using the PLS regression method with three factors (Figure 7), yielding a coefficient of determination and RMSEC of 0.24 and 0.55, respectively. Cross-validation yielded coefficient of determination and RMSECV values of 0.56 and 0.22, respectively (Table 1). External validation was performed on 131 samples not used to calibrate the model, generating a coefficient of determination of 0.10 and an RMSEP of 0.54. This model showed low predictive capacity, there is low range between the laboratory data (glucose content ranging from 44.5 to 48.5%), the variability was low in this trait. However, the model for the prediction of glucose content must not be used for screening application.

The xylan model was calibrated using the PLS regression method with six factors (Figure 7), yielding a coefficient of determination and RMSEC of 0.61 and 0.57, respectively. Cross-validation yielded coefficient of determination and RMSECV values of 0.57 and 0.60, respectively (Table 1). External validation was performed on 177 samples not used to calibrate the model, generating a coefficient of determination of 0.36 and an RMSEP of 0.68.

The results are very much lower than others in literature. VIANA et al. (2010) built a calibration model for glucose content in *Eucalyptus* spp., finding coefficient of determination of 0.44 and 0.86 and errors ranging from 0.78 to 1.15 in cross-validation. For xylan calibration models had coefficient of determination 0.51 and 0.90 and errors ranging from 0.12 to 0.37 in cross-validation. MILAGRES (2013) found coefficient of determination 0.59 and 0.64 and errors ranging from 0.58 to 0.92 for xylan and mananas, respectively.

On another hand, when comparing all parameters and RPD (Table 1) we assumed it is better to use the total carbohydrates model than isolated sugars. The total

carbohydrates model was calibrated using the PLS regression method with six factors (Figure 7), yielding a coefficient of determination and RMSEC of 0.65 and 0.66, respectively. Cross-validation yielded coefficient of determination and RMSECV values of 0.69 and 0.61, respectively (Table 1). External validation was performed on 131 samples not used to calibrate the model, generating a coefficient of determination of 0.53 and an RMSEP of 0.76.

The values reported in this study are statistically acceptable; however, a literature search did not return models for the total carbohydrates, only isolated values for each sugar showing that the values found are within the normal range. VIANA et al. (2010) reported a calibration model for xylan content in *Eucalyptus* spp., finding correlation coefficients between 0.72 and 0.95 and errors ranging from 0.12 to 0.37 in cross-validation. In *E. globulus*, NIR analysis gives good calibrations for predicting cellulose content. The standard errors of calibration and prediction were generally 1 or lower indicating that cellulose content can be predicted from NIR spectra (RAYMOND; SCHIMLECK, 2002).

3.3 GENETICS PARAMETERS BY NIR MODELS

NIR predicted values for wood basic density, lignin, extractive and total carbohydrate were determinate with low error (Figure 7). The individual additive genetic coefficient of variation (CV_{gi}) varied from 2.3 to 16.5 and the average heritability (h^2_{mp}) among wood traits in progenies were 0.72 and 0.81 showing medium to high magnitude values (COSTA et al., 2016), which shows that gains can be obtained with the selection for wood traits (Table 2). KIEN et al. (2008), evaluated cellulose content from breast-height increment cores in *E. urophylla*, found to moderate heritability for cellulose content (0.50), while the coefficient of additive genetic

variation for cellulose content was relatively low at 3.9%, for wood basic density heritability was (0.48) and coefficient of additive genetic variation 5.6%.

Table 2 - Genetics parameters estimated for wood traits determinate by NIR models.

GP/WT	Basic Density	Extractives	Lignin	Carbohydrates
CVgi	6.5	16.5	1.5	2.3
h ² mp	0.80	0.81	0.78	0.72

GP – genetic parameters; WT-wood traits;

The physicochemical properties of the wood were determined as presented in Table 3. The values found for total lignin (27.9%), extractives (1.5%) and wood basic density (0.395 g/cm³), were higher than *Eucalyptus urophylla* x *E. grandis* at 3 years old, according to Sousa (2008).

Table 3 – Wood physicochemical characterization of 4 years old *Eucalyptus benthamii* population in Laboratory.

Traits	Average	LI	LS	Error	CV (%)
WBD (g/cm ³)	0,388	0,30	0,50	0,03	8,1
Carbohydrates (%)	61,74	57,70	66,40	1,15	1,9
Extractives (%)	2,46	0,60	5,31	0,78	31,5
Klason Lignin, % (Insolúvel)	27,07	21,50	31,20	1,16	4,3
Klason Lignin, % (Solúvel)	4,43	2,70	5,80	0,40	8,9
Klason Lignin, % (Total)	31,50	26,80	35,50	1,15	3,6

4. CONCLUDING REMARKS

The NIR technology presents itself as a potential tool for *E. benthamii* breeding, enabling the early, non-destructive evaluation of important chemical wood features for the pulp production process and allowing cost reduction of chemical analysis, which may also be optimized.

The models for estimating lignin and extractive contents presented acceptable predictive ability indicating that the estimates presented small errors and can be applied in breeding programmes of *Eucalyptus benthamii*. However, the statistics associated to the models for glucose and xylan is not satisfactory for initial screenings.

The model for wood basic density exhibited a low error associated with this technology (despite its low correlation). The model for wood basic density is unsuitable to estimate the values of unknown samples.

5. REFERENCES

- ALVES, I.C.N.; GOMIDE, J.L.; COLODETTE, J.L. and SILVA, H.D. Technological characterization of *Eucalyptus benthamii* wood for Kraft pulp production. **Ciência Florestal**, v. 21, p. 167-174, 2011.
- CASTILLO, R.; CONTRERAS, D.; FREERA, J.; RUIZA, J. and VALENZUELA, S. Supervised pattern recognition techniques for classification of *Eucalyptus* species from leaves NIR spectra. **Journal of Chilean** v. 53, p. 1709-1713, 2008.
- COSTA, R.M.L.; ESTOPA, R.A.; BIERNASKI, F.A.; MORI, E.S. Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cambage por diferentes métodos de seleção. **Scientia Forestalis**, v. 44, n. 109, p. 105-113, 2016.
- DAVRIEUX, F.; ROUSSET, P.L.A.; PASTORE, M.C.T.; MACEDO, L.A. and QUIRINO, W.F. Discrimination of native wood charcoal by infrared spectroscopy. **Química Nova**, v. 33, p. 1093-1097, 2010.
- ESPINOZA, J. A.; HODGE, G.R.; DVORAK W.S. The potential use of Near Infrared Spectroscopy to discriminate between pine species and their hybrids. **Journal of Near Infrared Spectroscopy**, v. 18, p. 437-447, 2012.
- FANTUZZI NETO, H. **Qualidade da madeira de *Eucalyptus* para produção de celulose Kraft**. 2012. 105p. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.
- FONSECA, S. M.; RESENDE, M. D.V.; ALFENAS, A.C.; GUIMARAES, L.M.S.; ASSIS, T.F.; GRATTAPAGLIA, D. **Manual prático de melhoramento genético do *Eucalyptus***. Viçosa: Editora UFV, 2010. 200p.
- GREAVES; B.L.; SCHIMLECK, L.R.; BORRALHO, N.M.G. and MICHELL, A.J. Genetic control and repeatability of near infrared reflectance from *Eucalyptus nitens* woodmeal, **Appita J.** v. 49, p. 423-426, 1996.
- GOMIDE, J.L. and DEMUNER, B.J. Determination of lignin content in woody material: modification Klason Method. **O Papel**, v. 47, p. 8, 1986.

HEIN, P.R.G. and CHAIX, G. NIR spectral heritability: a promising tool for wood breeders. **Journal of Near Infrared Spectroscopy**, v. 22, p. 141-147, 2014.

HEIN, P.R.G.; LIMA, J.T.; CHAIX, G. Effects of sample preparation on NIR spectroscopic estimation of chemical properties of *Eucalyptus urophylla* S.T. Blake wood. **Holzforschung**, v. 64, p. 45-54, 2010.

HEIN, P.R.G.; LIMA, J.T. and CHAIX, G. Robustness of models based on near infrared spectra to predict the basic density in *Eucalyptus urophylla* wood. **J. Near Infrared Spectroscopy**, v. 17, p. 141-150, 2009.

HOFFMEYER, P. and Pedersen, J.G. Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy. **Holz als Roh und Werkstoff**, v. 53, p. 3, 1995.

IBÁ. Brazilian tree industry, *Statistical yearbook 2016*, Base year 2015.

KIEN, N.D.; QUANG, T.H.; JANSSON, G.; HARWOOD, C.; CLAPHAM, D. and VON ARNOLD, S. "Cellulose content as a selection trait in breeding for Kraft pulp yield in *Eucalyptus urophylla*", **Ann. Forest Sci.** v. 66, p. 708, 1-8, (2009).

KENNARD, R.W. and STONE, L.A. **Computer aided design of experiments**. Technometrics, v. 11, p. 137-148, 1969.

LAZZAROTTO, M. and MAGALHÃES, W.L.E. Near infrared spectroscopy use for non-destructive prediction in *Eucalyptus benthamii* e *Eucalyptus pellita* wood basic density. **Technical Notice - Embrapa Florestas**, n. 342, 2014.

MAESTRI, R. **Modelo de crescimento e produção para povoamentos clonais de *Eucalyptus grandis* considerando variáveis ambientais**. 2003. 143 p. PhD thesis. Universidade Federal do Paraná, Curitiba.

MILAGRES, F. R.; GOMIDE, J. L.; MAGATON, A. S.; FANTUZZI NETO, H. The influence of the age on near infrared spectroscopy models to prediction of the *Eucalyptus* spp. wood properties. **Revista Árvore** v. 37, p. 1165-1173, 2013.

MILAGRES, F.R. **Espectroscopia DE infravermelho próximo para predição de propriedade da madeira e híbridos de *Eucalyptus* spp.** 2013. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.

NISGOSKI, S.; MUÑIZ, G.I.B.; MORRONE, S.R.; SCHARDOSIN, F.Z.; FRANÇA, R.F. NIR and anatomy of wood and charcoal from Moraceae and Euphorbiaceae species. **Ciência da Madeira** v. 6, n. 3, p. 183-190, 2015.

OLIVEIRA, A.A; SIQUEIRA, P.H.; NISGOSKI, S.. BOLZON DE MUNIZ, G.I.; FERREIRA, J.H. Identificação de Madeiras utilizando a Espectrometria no Infravermelho Próximo e Redes Neurais Artificiais. **Tendências em Matemática Aplicada e Computacional**, v. 16, n. 2 p.81-95, 2015.

PASQUINI, C.; CARNEIRO, J.G.C. and BOMFIM, P.M. Development and validation of multivariate model employing Near Infrared Spectroscopy for *Eucalyptus* wood characteristics estimations. **O Papel**, v. 68, p. 3, 2007.

RAMADEVI, P.; HEGDE, D.V.; VARGHESE,M; KAMALAKANNAN, R, GANAPATHY, S.P. and GURUMURTHY, D.S. Evaluation of lignin syringyl/guaiacyl ratio in *Eucalyptus camaldulensis* across three diverse sites based on near infrared spectroscopic calibration modelling with five *Eucalyptus* species and its impact on Kraft pulp yield. **J. Near Infrared Spectroscopy**, v. 24, p. 529-536, 2016.

RAYMOND, C.A. and SCHIMLECK, L.R. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. **Canadian Journal of Forest Research**, v. 32, p. 170–176, 2002.

RESENDE, M.D.V. Selegen-REML/BLUP : Sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Colombo: Embrapa Florestas, 2007. 359p.

ROSSO, S.; MUNIZ, G.I.B; MATOS, J.L.M.; HASELEIN, C.R.; HEIN, P.R.G, LOPES, M.C. Density of *Eucalyptus Grandis* W. Hill Ex Maiden using Near Infrared Spectroscopy. **Cerne**, Lavras, v. 19, n. 4, p. 647-652, 2013.

SAMISTRARO, G.; DE MUNIZ, G; G. I. B.; PERALTA-ZAMORA, P.; CORDEIRO, G.A. Previsão das propriedades físicas do papel Kraft por espectroscopia no infravermelho próximo (NIR) e regressão por mínimos quadrados parciais (PLS). **Química Nova**, v.32, n. 6, p.1422-1425, 2009.

SAVITZKY, A. and GOLAY, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. **Analytical Chemistry**, v. 36, p. 8, 1997.

SCHIMLECK, L.R.; PAYNE, P. and WEARNE, R.H. Determination of important pulp properties of hybrid poplar by near infrared spectroscopy. **Wood and Fiber Science**, v. 37, p. 3, 2005.

SCHIMLECK, L.R.; MORA, C.; DANIELS, R.F. Estimation of physical wood properties of green *Pinus taeda* radial samples by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 33, p. 2297-2305, 2003.

SOUSA, L.C. **Use of Near Infra Red Spectroscopy to predict *Eucalyptus* wood characteristics for cellulose production.** 2008. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.

SOUSA, L.C. GOMIDE, J.L.; MILAGRES, F.R. ALMEIDA, D.P. Development of NIRS calibration models for minimization of *Eucalyptus* spp wood analysis. **Ciência Florestal**, Santa Maria, v. 21, n. 3, p. 591-599, 2011.

SANTOS, R.B.; GOMIDE, J.L.; SOUSA, L.C. Prediction of wood and Kraft Pulp quality by Near Infrared Spectroscopy. **R. Árvore**, Viçosa-MG, v.33, n.4, p.759-767, 2009.

Technical Association of the Pulp and Paper Industry-TAPPI. T280 pm-99 standard - Acetone extractives of wood and pulp. **TAPPI Press**, 2000.

VIANA, L.C.; TRUGILHO, P.F.; HEIN, P.R.G.; SILVA, J.M. and LIMA, J.T. Calibration models and near infrared spectroscopy for predicting chemical properties and wood basic density in *Eucalyptus*. **Ciência Florestal**, v. 20, p. 2, 2010.

WALLIS, A.F.A.; WEARNE, R.H. and WRIGHT, P.J. Analytical characteristics of plantation eucalypt woods relating to Kraft pulp yields. **Appita Journal** v. 49, p. 6, 1996.

WILLIAMS, P.C.; SOBERING, D.C. Comparison of commercial Near Infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, v. 1, p. 25-33, 1993.

6. CONSIDERAÇÕES FINAIS

O NIR e a Seleção Genômica são biotecnologias que podem acelerar o programa de melhoramento do *Eucalyptus benthamii* e que se confirmaram promissoras nesse estudo, validações externas serão necessárias para comprovar as capacidades preditivas, principalmente da Seleção Genômica.

Pôde-se constatar que com um número menor de marcadores, em torno de 2 a 3 mil, é possível acessar a variabilidade da população, para todas as características estudadas. Porém, como existe um chip de 60k, é melhor realizar as genotipagens com o chip completo e depois por meio de análises definirem a melhor estratégia de predição.

A seleção superprecoce via seleção genômica pode aumentar a eficiência da seleção, se realizada ainda na fase de progênie do programa de melhoramento, para as características de volume e qualidade da madeira.

A tecnologia NIR também se apresenta como potencial ferramenta para o programa de melhoramento genético da espécie, tornando-se possível a avaliação precoce e não destrutiva de características químicas importantes para o processo de produção de polpa celulósica, além da redução de custos com análises químicas, que poderão ser otimizadas.

O próximo passo desse trabalho será realizar validações externas e independentes, visando avaliar de forma mais precisa as capacidades preditivas. Também será realizada a genotipagem de indivíduos relacionados com essa população em estudo, visando selecionar indivíduos por meio do valor genético genômico.

As dificuldades encontradas neste estudo se referem à coleta de amostras de madeira com o trado de incremento. Por mais de uma vez o trado quebrou e uma nova compra foi realizada. Isso provocou demora e intervalos longos de tempo entre a coleta da primeira e última amostra.

7. REFERÊNCIAS GERAIS

ALVES, I.C.N.; GOMIDE, J.L.; COLODETTE, J.L. and SILVA, H.D. Technological characterization of *Eucalyptus benthamii* wood for Kraft pulp production. **Ciência Florestal**, v. 21, p. 167-174, 2011.

AZEVEDO, C.F.; RESENDE, M.D.V.; SILVA, F.F.; VIANA, J.M.S.; VALENTE, M.S.F.; RESENDE JR, M.F.R.; OLIVEIRA, E.J. New accuracy estimators for genomic selection with application in a cassava (*Manihot esculenta*) breeding program. **Genetics and Molecular Research**, v. 15, p. 1, 2016.

BERMINGHAM, M.L.; PONG-WONG, R., SPILIOPOULOU, A.; HAYWARD, C.; RUDAN, I.; CAMPBELL H.; WRIGHT, A.F.; WILSON, J.F.; AGAKOV, F.; NAVARRO, P.; HALEY, C.S. Application of high-dimensional feature selection: evaluation for genomic prediction in man. **Scientific Reports**. may, 2015. DOI: 10.1038/srep10312

BERNARDO, R; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, n.3, p.1082-1090, 2007.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, 2008.

BUTCHER, P.A., SKINNER, A.K., GARDINER, C.A. Increased inbreeding and inter-species gene flow in remnant populations of the rare *Eucalyptus benthamii*. **Conservation Genetics**, v. 6, n. 2, p.213–226, 2005.

CASTILLO, R.; CONTRERAS, D.; FREERA, J.; RUIZA, J. and VALENZUELA, S. Supervised pattern recognition techniques for classification of *Eucalyptus* species from leaves NIR spectra. **Journal of Chilean** v. 53, p. 1709-1713, 2008.

COSTA, R.M.L.; ESTOPA, R.A.; BIERNASKI, F.A.; MORI, E.S. Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cambage por diferentes métodos de seleção. **Scientia Forestalis**, v. 44, n. 109, p. 105-113, 2016.

COSTA, R.M.L.; ESTOPA, R.A.; BIERNASKI, F.A.; MORI, E.S. Predição de ganhos genéticos em progênies de *Eucalyptus benthamii* Maiden & Cambage por diferentes métodos de seleção. **Scientia Forestalis**, v. 45, n. 113, p. 105-113, 2017.

CROSSA J.; DE LOS CAMPOS, G.; PEREZ, P. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v.186, n.2, p.713-724, 2010.

DAVRIEUX, F.; ROUSSET, P.L.A.; PASTORE, M.C.T.; MACEDO, L.A. and QUIRINO, W.F. Discrimination of native wood charcoal by infrared spectroscopy. **Química Nova**, v. 33, p. 1093-1097, 2010.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v.193, p.327-345, 2013.

DEKKERS, J.C.M. Commercial application of marker and gene assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, n.13, p.313-328, 2004.

DEKKERS, J.C.M. Marker-assisted selection for commercial crossbred performance. **Journal of Animal Science**, v. 85. p.2104-2114, 2007.

DEKKERS, J. C. M. Application of Genomics Tools to Animal Breeding. **Current Genomics**, v.13, p. 207-212, 2012.

EISBERG, R.; RESNICK, R. **Física quântica**: átomos, moléculas, sólidos, núcleos e partículas. 4º ed. Rio de Janeiro: Campus, 372p, 1986.

ESPINOZA, J. A.; HODGE, G.R.; DVORAK W.S. The potential use of Near Infrared Spectroscopy to discriminate between pine spices and their hybrids. **Journal of Near Infrared Spectroscopy**, v. 18, p. 437-447, 2012.

FANTUZZI NETO, H. **Qualidade da madeira de *Eucalyptus* para produção de celulose Kraft**. 2012. 105p. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.

FONSECA, S. M.; RESENDE, M. D.V.; ALFENAS, A.C.; GUIMARAES, L.M.S.; ASSIS, T.F.; GRATTAPAGLIA, D. **Manual prático de melhoramento genético do *Eucalyptus***. Viçosa: Editora UFV, 2010. 200p.

GARCIA, C.C. Avaliação de modelos preditivos de seleção genômica ampla em testes de progênies e testes clonais de *Eucalyptus*. 2016, 92p. **Dissertação** (Mestrado em Ciências: Recursos Florestais) – Escola Superior de Agricultura/USP, Piracicaba.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and Rapd markers. **Genetics**, Bethesda, v. 137, n. 4, p. 1121-1137, 1994.

GRATTAPAGLIA D; RESENDE M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. (Ed.). **Genomics of plant genetic resources**. Dordrecht: Springerchap. 26, p. 651-682, 2014.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 323-330, 2007.

GOLDSCHMID, O. Ultraviolet spectra. In: Sarkanen, K.V. e Ludwig, C.H. Ed. **Lignins**. New York: Wiley-Interscience, 1971. p.241-266.

GOMIDE, J.L. and DEMUNER, B.J. Determination of lignin content in woody material: modification Klason Method. **O Papel**, v. 47, p. 8, 1986.

GORE, M.A.; CHIA J.M.; ELSHIRE, R.J. A First-Generation Haplotype Map of Maize. **Science**, v. 326, n. 5956, p. 1115-1117, 2009.

GREAVES; B.L.; SCHIMLECK, L.R.; BORRALHO, N.M.G. and MICHELL, A.J. Genetic control and repeatability of near infrared reflectance from *Eucalyptus nitens* woodmeal, **Appita J.** v. 49, p. 423-426, 1996.

HAYES, B.J.; VISSCHER, P.M.; GODDARD, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, v. 91, n. 1, p. 47-60, 2009.

HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J.L. Genomic selection for crop improvement. **Crop Science**, v. 49, n. 1, p.1-12, 2009.

HEIN, P.R.G.; LIMA, J.T.; Chaix, G. "Robustness of models based on near infrared spectra to predict the basic density in *Eucalyptus urophylla* wood". **Journal of. Near Infrared Spectroscopy**, v.17, p.141-150, 2009.

HEIN, P.R.G.; LIMA, J.T.; CHAIX, G. Effects of sample preparation on NIR spectroscopic estimation of chemical properties of *Eucalyptus urophylla* S.T. Blake wood. **Holzforschung**, v. 64, p. 45-54, 2010.

HEIN, P.R.G. and CHAIX, G. NIR spectral heritability: a promising tool for wood breeders. **Journal of Near Infrared Spectroscopy**, v. 22, p. 141-147, 2014.

HENRIQUES, E.P; MORAES, C.B; SEBBENN, A.M.; TOMAZELLO FILHO, M.; MORAES, M.L.T; MORI, E.S. Estimativa de parâmetros genéticos para caracteres silviculturais e densidade do lenho em teste de progênies de *Eucalyptus urophylla*. **Scientia Forestalis**, v. 44, n. 109, p. 105-113, 2016.

HOFFMEYER, P. and Pedersen, J.G. Evaluation of density and strength of Norway spruce by near infrared reflectance spectroscopy. **Holz als Roh und Werkstoff**, v. 53, p. 3, 1995.

IBÁ. **Indústria Brasileira de Árvores**, ano base 2015, Brasília, 2016.

IWATA, H.; HAYASHI, T.; TSUMURA, Y. Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. **Tree Genetics & Genomes**, v.7, n.4, p. 747-758, 2011.

JUNGHANS, D.T.; ALFENAS, A.C., BROMMONSCHENKEL, S.H., ODA, S., MELLO, E.J., GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of major gene with RAPD markers. **Theoretical and Applied Genetics**, v. 108, n. 1, p. 175-180, 2003.

KIEN, N.D.; QUANG, T.H.; JANSSEN, G.; HARWOOD, C.; CLAPHAM, D. and VON ARNOLD, S. "Cellulose content as a selection trait in breeding for Kraft pulp yield in *Eucalyptus urophylla*", **Ann. Forest Sci.** v. 66, p. 708, 1-8, (2009).

KENNARD, R.W. and STONE, L.A. **Computer aided design of experiments**. Technometrics, v. 11, p. 137-148, 1969.

LAZZAROTTO, M. and MAGALHÃES, W.L.E. Near infrared spectroscopy use for non-destructive prediction in *Eucalyptus benthamii* e *Eucalyptus pellita* wood basic density. **Technical Notice - Embrapa Florestas**, n. 342, 2014.

LANDE, R.; THOMPSON, R. Efficiency of marker assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, n. 3, p. 743-756, 1990.

LEGARRA A., ROBERT-GRANIE C., MANFREDI E., ELSEN J. M. Performance of genomic selection in mice. **Genetics**, v. 180, p. 611-618, 2008.

LIMA, B.M. Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data. 2014, 92p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura/USP, Piracicaba.

LONG, N.; GIANOLA, D.; ROSA, G. J.M.; WEIGEL, K.A; AVENDAÑO, S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. **Journal of Animal Breeding and Genetics**, v. 124, n.6, p. 377-389, 2007.

MAESTRI, R. **Modelo de crescimento e produção para povoamentos clonais de *Eucalyptus grandis* considerando variáveis ambientais**. 2003. 143 p. PhD thesis. Universidade Federal do Paraná, Curitiba.

MEUWISSEN, T. H. E.; GODDARD, M. E.; HAYES, B. J. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p.1819-1829, 2001.

MEUWISSEN, T.H.E. Genomic selection: marker assisted selection on genome-wide scale. **Journal of Animal Breeding and Genetics**, v. 124, n. 6, p. 321-322, 2007.

MILAGRES, F.R. **Espectroscopia DE infravermelho próximo para predição de propriedade da madeira e híbridos de *Eucalyptus* spp.** 2013. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.

MILAGRES F. R.; GOMIDE , J. L.; MAGATON, A. S.; FANTUZZI NETO H. The influence of the age on near infrared spectroscopy models to prediction of the *Eucalyptus* spp. wood properties. **Revista Árvore**, v. 37, p.1165-1173, 2013.

MISSIAGIA, A. A.; PIACEZZI, A. L.; GRATTAPAGLIA, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. **Tree Genetics and Genomes**, Heidelberg, v. 1, n. 1, p. 79-84, Oct. 2005.

MUIR, W.M. Comparison of genomic and traditional BLUP estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, v. 124, n.6, p. 342-355, Dec. 2007.

NISGOSKI, S.; MUÑIZ, G.I.B.; MORRONE, S.R.; SCHARDOSIN, F.Z.; FRANÇA, R.F. NIR and anatomy of wood and charcoal from Moraceae and Euphorbiaceae species. **Ciência da Madeira** v. 6, n. 3, p. 183-190, 2015.

OLIVEIRA, A.A.; SIQUEIRA, P.H.; NISGOSKI, S.. BOLZON DE MUNIZ, G.I.; FERREIRA, J.H. Identificação de Madeiras utilizando a Espectrometria no Infravermelho Próximo e Redes Neurais Artificiais. **Tendências em Matemática Aplicada e Computacional**, v. 16, n. 2 p.81-95, 2015.

PASQUINI, C.; CARNEIRO, J.G.C. and BOMFIM, P.M. Development and validation of multivariate model employing Near Infrared Spectroscopy for *Eucalyptus* wood characteristics estimations. **O Papel**, v. 68, p. 3, 2007.

PRYCE, J.E.; DAETWYLER, H.D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. **Animal Production Science**, v.52, n. 3, p.107-114, Dec. 2012.

RAMADEVI, P.; HEGDE, D.V.; VARGHESE,M; KAMALAKANNAN, R, GANAPA-THY, S.P. and GURUMURTHY, D.S. Evaluation of lignin syringyl/guaiacyl ratio in *Eucalyptus camaldulensis* across three diverse sites based on near infrared spec-troscopic calibration modelling with five *Eucalyptus* species and its impact on Kraft pulp yield. **J. Near Infrared Spectroscopy**, v. 24, p. 529-536, 2016.

RAYMOND, C.A.; SCHIMLECK, L.R. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. **Canadian Journal of Forest Research**, v. 32, p. 170–176, 2002.

RESENDE, M.D.V.; BARBOSA, M.H.P. Melhoramento genético de plantas de propagação assexuada. Colombo: **Embrapa Florestas**, 2005. 130p.

RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RESENDE, M.D.V. Selegen-REML/BLUP: Sistema estatístico e seleção genética computadorizada via modelos lineares mistos. Colombo: **Embrapa Florestas**, 2007b. 359p.

RESENDE, M.D.V.; ASSIS, T.F. Seleção recorrente recíproca entre populações sintéticas multi-espécies (SRR-PSME) de *Eucalyptus*. **Pesquisa Florestal Brasileira**, v.57, p.57-60, 2008.

RESENDE, M.D.V.; LOPES, P.S.; SILVA, R.L; PIRES, I.E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal**, n. 56, 2008.

RESENDE, M.F.JR; MUNOZ, P.; RESENDE, M.D.V.; GARRICK, D. J.; FERNANDO, R. L.; DAVIS, J. M; JOKELA, E. J.; MARTIN,T. A.; PETER, G. F.; KIRST, M. Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.). **Genetics**, v.190, n.4, p. 1503-1510, 2012a.

RESENDE, M.D.V.; RESENDE, M.F.R.; SANSALONI, C.P. MISSIAGGIA AA, AGUIAR A.M., ABAD J.M., TAKAHASHI E.K., ROSADO A.M., FARIA D.A. et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, v. 194, n.1, p.116-128, 2012.

RESENDE, M.D.V. DE; SILVA, F. F. E. ; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional**: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência. 1. ed. Visconde do Rio Branco: Suprema, 2014, v. 1. 881p.

REZENDE, B.A. seleção genômica ampla para volume e qualidade da madeira em *Eucalyptus*. 2015, 74p. **Tese** (Doutorado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras.

ROSADO, A.M.; ROSADO, T.B.; RESENDE JUNIOR, M.F.R.; BHERING, L.L.; CRUZ, C.D. Ganhos genéticos preditos por diferentes métodos de seleção em progênies de *Eucalyptus urophylla*. **Pesquisa Agropecuária Brasileira**, v.44, n.12, p.1653-1659, 2009.

ROSS, R.J.; BRASHAW, B.K.; PELLERIN, R.F. Nondestructive evaluation of wood. **Forest Products Journal**, v. 48, n. 1, p.14-19, 1998.

ROSSO, S.; MUNIZ, G.I.B; MATOS, J.L.M.; HASELEIN, C.R.; HEIN, P.R.G, LOPES, M.C. Density of *Eucalyptus grandis* W. Hill Ex Maiden using Near Infrared Spectroscopy. **Cerne**, Lavras, v. 19, n. 4, p. 647-652, 2013.

SAMISTRARO, G.; DE MUNIZ, G; G. I. B.; PERALTA-ZAMORA, P.; CORDEIRO, G.A. Previsão das propriedades físicas do papel Kraft por espectroscopia no infravermelho próximo (NIR) e regressão por mínimos quadrados parciais (PLS). **Química Nova**, v.32, n. 6, p.1422-1425, 2009.

SANTOS, R.B.; GOMIDE, J.L.; SOUSA, L.C. Prediction of wood and Kraft Pulp quality by Near Infrared Spectroscopy. **R. Árvore**, Viçosa-MG, v.33, n.4, p.759-767, 2009.

SAVITZKY, A. and GOLAY, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. **Analytical Chemistry**, v. 36, p. 8, 1997.

SCHAEFFER, L.R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, v. 123, n.4, p. 218-223, 2006.

SCHIMLECK, L.R.; MORA, C.; DANIELS, R.F. Estimation of physical wood properties of green *Pinus taeda* radial samples by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 33, p. 2297-2305, 2003.

SCHIMLECK, L.R., PAYNE, P.; WEARNE, R.H. Determination of important pulp properties of hybrid poplar by near infrared spectroscopy. **Wood and Fiber Science**, v.37, n. 3, p. 462-471, 2005.

SILVA-JUNIOR, O.B.; GRATTAPAGLIA, D. Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. **New Phytologist**, London, v. 208, n. 3, p. 830-845, 2015.

SOLBERG, T.R.; SONESSON, A.; WOOLIAMs, J.; MEUWISSEN, T.H.E. Genomic selection using different marker types and density. **Journal of Animal Science**, v.86, n.10, 2447-2454, 2008.

SOUSA, L.C. **Use of Near Infra Red Spectroscopy to predict *Eucalyptus* wood characteristics for cellulose production.** 2008. PhD thesis. Universidade Federal de Viçosa, Viçosa/MG.

SOUSA, L.C.; SANTOS, R. B.; GOMIDE, J.L. Prediction of wood and kraft pulp quality by Near Infrared Spectroscopy (NIR). **Revista Árvore**. v.33, p. 759-767, 2009.

SOUSA, L.C. GOMIDE, J.L.; MILAGRES, F.R. ALMEIDA, D.P. Development of NIRS calibration models for minimization of *Eucalyptus* spp wood analysis. **Ciência Florestal**, Santa Maria, v. 21, n. 3, p. 591-599, 2011.

SU, G; BRONDUM, RF; MA, P; GULDBRANDTSEN, B.; AAMAND, GR; LUND, MS. Comparison of genomic predictions using medium-density (similar to 54,000) and high-density (similar to 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. **J Dairy Sci**. v. 95, n.8, p.4657-4665, 2012.

Technical Association of the Pulp and Paper Industry-TAPPI. T280 pm-99 standard - Acetone extractives of wood and pulp. **TAPPI Press**, 2000.

THAMARUS, K.; GROOM, K.; BRADLEY, A.; RAYMOND, C.A.; SCHIMLECK, L.R.; WILLIAMS, E.R.; MORAN, G.F. Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. **Theor Appl Genet** v. 109, p. 856-864, 2004.

VIANA, L.C.; TRUGILHO, P.F.; HEIN, P.R.G.; SILVA, J.M; LIMA, J.T. Calibration models and near infrared spectroscopy for predicting chemical properties and wood basic density in *Eucalyptus*. **Ciência Florestal**, v. 20, p. 2, 2010.

VIANA, A. P. ; RESENDE, M.D.V. DE . Genética Quantitativa no Melhoramento de Fruteiras. 1. ed. Rio de Janeiro: **Interciência**, v. 1. 284p, 2014.

ZHANG, Z., DING X, LIU J., ZHANG, Q., DE KONING, D.J. Accuracy of genomic prediction using 778 low-density marker panels. **J Dairy Sci.** v. 94, n.7, p. 3642-3650, 2011.

ZHONG, S.;DEKKERS, J.C.M.;FERNANDO, R.L.;JANNINK, J-L. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, n.1, p. 355-364, 2009.

WALLIS, A.F.A.; WEARNE, R.H. and WRIGHT, P.J. Analytical characteristics of plantation eucalypt woods relating to Kraft pulp yields. **Appita Journal** v. 49, p. 6, 1996.

WILLIAMS, P.C.; SOBERING, D.C. Comparison of commercial Near Infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, v. 1, p. 25-33, 1993.

WRAY,N.R.; YANG, J.; HAYES, B.J.; PRICE, A.L.; GODDARD, M. E.; VISSCHER, P.M. Pitfalls of predicting complex traits from SNPs. **Nature Rev Genet.** v. 14, p.507–515. 2013.

WONG, C.K.; BERNARDO, R. Genome wide selection in oil palm: increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics**, v.116, n.6, p. 815-824, 2008.